Exploring the Epoch of Reionization with the 21 cm line

Dijana Vrbanec

München 2019

Exploring the Epoch of Reionization with the 21 cm line

Dijana Vrbanec

Dissertation
an der Ludwig-Maximilians-Universität München

vorgelegt von
Dijana Vrbanec
aus Zagreb, Croatia

München, den 21.06.2019

Erstgutachter: Simon White
Zweitgutachter: Jochen Weller
Tag der mündlichen Prüfung: 25.07.2019

Contents

List of Figures

List of Abbreviations

CMB - cosmic microwave background
EoR - Epoch of Reionization
EW - equivalent width
FoV - field of view
FM - frequency modulated radio band
GRB - gamma ray burst
HERA - Hydrogen Epoch of Reionization Array
HSC - Hyper Supreme Cam
IGM - intergalactic medium
IMF - initial mass function
ISM- interstellar medium
kSZ - kinetic Sunyaev-Zel'dovich effect
LAE - Lyman Alpha Emitter
LBG - Lyman Break Galaxy
LLS - Lyman Limit System
LOS - line of sight
LOFAR - Low Frequency Array
MWA - Murchison Widefield Array
NIRB - near-infrared background
PFS - Prime Focus Spectrograph
PSF - point spread function
Pop-III stars - Population-III stars
SEFD - system equivalent flux density
SFR - star formation rate
SN - supernova
QSO - quasi-stellar object
S/N - signal to noise
SKA - Square Kilometre Array
UV - ultraviolet
WFIRST - Wide Field Infrared Survey
WMAP - Wilkinson Microwave Anisotropy Probe

Zusammenfassung

Die Reionisierung ist der letzte große Phasenübergang des Universums, der den Übergang von einem völlig neutralen zu einem vollständig ionisierten intergalaktischen Medium markiert. Dieser Übergang wird von jungen sternbildenden Galaxien angetrieben, die ihr umgebendes Gas reionisieren. Heute konzentrieren sich erheblichen Beobachtungsanstrengungen auf die Beobachtung des 21cm-Signals des intergalaktischen neutralen Wasserstoffs, der reionisiert wird, und der Galaxien, die Reionisierung antreiben. Um diese Beobachtungen zu planen werden Vorhersagen über die Beobachtungen gemacht, um die besten Beobachtungsstrategien und Synergien zwischen verschiedenen Instrumenten zu finden. Mit dieser Arbeit schließe ich mich diesen Bemühungen an, indem ich mich auf drei besondere Probleme fokussiere. Der erste Fokus liegt auf Synergien zwischen laufenden Beobachtungen von LOFAR und HSC, die 21cm-Emission bzw. Lyman-α emittierende Galaxien beobachten werden, sowie auf Synergien zwischen SKA, einem Radioteleskop der nächsten Generation, das 21cm-Signal beobachtet, und HSC, PFS und WFIRST, die Galaxien beobachten. Ich zeige, dass die Kreuzkorrelation zwischen LOFAR- und HSC-Beobachtungen eine Anti-Korrelation auf den größten Skalen zeigen und damit beweisen sollten, dass das Signal, das LOFAR als Instrument der ersten Generation beobachtet, tatsächlich aus der Epoche der Reionisierung stammt. Kreuzkorrelationen zwischen SKA und HSC können diese Beobachtungen verbessern, und in Kombination mit der Synergie zwischen SKA und WFIRST könnte sich sogar die zeitliche Entwicklung dieser Kreuzkorrelation zeigen. Ein weiterer Fokus zeigt, dass SKA potenziell eindimensionale 21cm-Leistungsspektren beobachten könnte, die einen Einblick in kleinskalige kosmologische Strukturen geben würden, im Gegensatz zu den zwei- und dreidimensionalen Beobachtungen, die sich auf große Dimensionen konzentrieren. Ich untersuche auch die bestmöglichen Beobachtungsstrategien für die zweidimensionalen Beobachtungen von SKA, und zeige, dass die Beobachtung senkrecht zur Sichtlinie viel besser wäre als die Beobachtung in Sichtlinie aufgrund von um Größenordnungen kleineren Rauschens in Winkelrichtung als in Frequenzrichtung. Mein letzter Fokus liegt auf unserem theoretischen Verständnis des 21cm-Waldes. Ich zeige, dass das Feedback von Supernovae, sofern sie so implementiert werden, dass die Reionisationsgeschichte nicht verändert wird, keine statistischen Veränderungen in den Merkmalen der 21cm-Wald-Spektren im Vergleich zum Fall ohne Supernovae induzieren. Ich bestätige auch, dass die verschiedenen Auflösungen der Simulationen, die zur Erzeugung der 21cm-Wald-Spektren verwendet wurden, für den größten Anteil an den Verschiedenheiten in den simulierten Spektren zuständig sind.

Summary

The Epoch of Reionization is the last major phase transition of the Universe, marking the passage from a completely neutral to a fully ionized intergalactic medium. This transition is driven by young star-forming galaxies that are reionizing their surrounding gas. Today, substantial observational effort is focused on observing the 21cm signal of the intergalactic neutral hydrogen that is being reionized and the galaxies driving the reionization. To help guide this observational effort predictions of the possible observations are being made, to asses the best observational strategies and possible synergies between different instruments. With this thesis, I am joining these efforts by focusing on three particular problems. I focus on possible synergies between ongoing observations of LOFAR and HSC that are going to observe 21cm emission and Lyman-α emitting galaxies respectively, as well as synergies between the next-generation survey SKA, observing the 21cm signal, and HSC, PFS, and WFIRST observing the galaxies. I show that cross-correlation between LOFAR and HSC observations should show an anti-correlation on the largest scales and thus prove that the signal LOFAR is observing, as a first-generation instrument, is indeed from the Epoch of Reionization. Cross-correlations between SKA and HSC can improve on this effort and combined with the synergy between SKA and WFIRST even show the evolution of this cross-correlation. I also focus on showing that SKA could potentially observe 21cm line-of-sight one-dimensional power spectra, which would give insight on cosmological small scales. Unlike the two- and three- dimensional observations which focus on large scales. I also examined the best possible observational strategies for SKA's two-dimensional observations, showing that observing perpendicular to line-of-sight would be much better than observing in direction of line-of-sight due to orders of magnitude smaller noise in angular direction than in frequency direction. Lastly, I focus on our theoretical understanding of the 21cm forest. I show that when it comes to supernovae feedback, when implemented in such a way not to change the reionization history as opposed to without implementing it, no statistical changes are found in the features of the 21cm forest spectra. I also confirmed that the changes in the resolution of simulations used to build the 21 cm forest spectra induced the most changes in the spectra. I show that LOFAR could observe the most prominent features in a 21cm forest spectra from likes of Cygnus A high-redshift radio loud source, with a signal to noise ratio $S/N \sim 20$. SKA, instead, could observe even the smallest features, with $S/N \sim 200$.

Chapter 1

Introduction

The Epoch of Reionization is a period in the evolution of the Universe approximately 500 million to 1 billion years after the Big Bang, where intergalactic hydrogen changes from being completely neutral to being completely ionized. It is an early stage in the evolution of galaxies, where galaxies are still young and small and tend to have strong star formation. That makes them perfect candidates for releasing copious amounts of ionizing radiation into the surrounding intergalactic medium, re-ionizing the Universe. As such, the Epoch of Reionization marks an important milestone in the evolution of the Universe - the transition from a neutral to an ionized Universe.

Today, we are at the forefront of observing the Epoch of Reionization, which will offer insight into the evolution and topology of the reionization process, sources of ionizing radiation, and the evolution of the galaxies during this period. The observational efforts made towards the detection of the Epoch of Reionization are mainly in two fields - observing the galaxies or observing the 21 cm signal from intergalactic neutral hydrogen. Hubble Space Telescope, with Ultra Deep Field and its subsequent deep field observations, Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey, Large Area Lyman Alpha survey, The Deep and Wide Narrowband survey, and Subaru started the monumental task of mapping high-redshift galaxies, and space facility Wide Field Infrared Survey Telescope (WFIRST) is being planned to continue these efforts. Low Frequency Array (LOFAR), Murchison Widefield Array (MWA), and Hydrogen Epoch of Reionization Array (HERA) have started to map the intensity of the 21 cm line from intergalactic neutral hydrogen, whose topology at different redshifts could give us great insights into the evolution of the Epoch of Reionization, sources of ionizing radiation, and the sizes of ionized regions at different redshifts, as will Square Kilometre Array (SKA) which is in construction.

Simultaneously, the advance in computational power paved the way to simulations of the evolution of the Universe and the Epoch of Reionization covering ever greater scales and dynamical range. Today, simulations with sizes of a 21cm survey's field of view, $\sim 500\ h^{-1}$ Mpc, are available and have great dynamical range, which is necessary to plan and test viability for the present and upcoming observational efforts, as well as to further our understanding of the small- and large- scale (astro-)physics governing the evolution of the Universe and the reionization process. By means of these simulations, astrophysicists

are finally being able to make well-defined predictions for a variety of observations.

As a part of the LOFAR collaboration I am joining these efforts of making predictions for observations with LOFAR and SKA at low radio frequencies, as well as combining these observations with galaxy observations to further the possible information extraction from these observations. In more details, in the subsequent chapters I will explore the following projects:

I explore the possible synergy between the upcoming observations of the 21 cm emission with LOFAR and Lyman-α Emitters with Subaru Hyper Supreme Cam [119] since both will observe the ELAIS-N1 field at redshift $z = 6.6$, as well as possible synergies between future measurements of the 21cm emission with the SKA and future galaxy surveys such as WFIRST and Subaru Prime Focus Spectrograph. I designed mock observations that have characteristics of the observations that will be obtained with these instruments and then I cross-correlated 21cm emission and Lyman-α Emitters mock observations from different pairs of instruments to test whether it would be possible to observe such a cross-correlation and to see what can be learned from it. I show that LOFAR and Subaru Hyper Supreme Cam should be able to observe an anti-correlation at the largest scales and that observations with SKA will have much smaller noise levels than LOFAR which will greatly increase the accuracy of its observations and its cross-correlations. For more information see Chapter 2.

In Chapter 3 I explore how the direction of observation does influence the two-dimensional 21cm power spectra observations with SKA. I also explore the possibility of probing the small scales with one-dimensional line-of-sight 21cm power spectra and asses their observability with SKA at different redshifts throughout the reionization epoch. I show that the two-dimensional 21cm power spectra observations with SKA are much more precise when measured perpendicular to the line-of-sight. It is also possible to probe the small scales with one-dimensional line-of-sight 21cm power spectra as long as it goes through a mostly neutral medium with sky averaged neutral hydrogen fraction ≥ 0.68.

I analyze the effectiveness of the 21cm forest as a probe of the galactic feedback on small-scales in Chapter 4. To do that I use radiation-hydrodynamical simulations that have proper modeling of radiative feedback which includes modelling for the injection of mass and energy by both core-collapse and Type Ia supernovae, and the mass loss due to winds from massive stars and asymptotic giant branch stars. I explore the influence of the supernovae feedback and the resolution effects on the 21cm forest. I show that simulations with supernovae feedback prescriptions implemented in such a way to follow the same reionization histories as in the simulations that do not have supernovae feedback included show minimal statistical differences in the various absorption features. The resolution, on the other hand, has a major influence on the small scales probed by the 21cm forest. It greatly affects the strength of the observed spectra and simulations with higher resolution have more of stronger absorption features that could be easily observed with LOFAR and even better with SKA.

The general conclusion of the thesis can be found in the Conclusion.

1.1 Epoch of Reionization

After the Big Bang the Universe was very hot and consisted of only simple baryonic and leptonic particles such as protons, neutrons and electrons. As the Universe grew it cooled down and the protons and neutrons started combining into ionized atoms of hydrogen, some helium and trace amounts of other elements. With further cooling ionized atoms started bounding electrons and turning into neutral atoms in the process called Recombination. This process happened around 300 000 years after the Big Bang (redshift $z \approx 1100$) and resulted in the decoupling of radiation from matter. We can observe this relic radiation today as cosmic microwave background (CMB) radiation. The process of recombination was followed by the period of the Dark Ages when the Universe remained neutral and there were no sources of light. During the Dark Ages dark matter particles began combining into dark matter haloes, gravitationally attracting baryons, which eventually condensed within the dark matter haloes and formed galaxies. Towards the end of the Dark Ages, 150 million years after Big Bang ($z \approx 30$), the first stars formed shedding light into the dark Universe and starting the Cosmic Dawn. The first stars heated the surrounding intergalactic medium (IGM) forming near-infrared and non-ionizing UV and X-ray backgrounds which raised the temperature of the gas in the Universe, and started the reionization process where the ionizing radiation escapes the galaxies and ionizes their surrounding IGM. The bulk of the reionization process happens around 500 million - 1 billion years after Big Bang ($z \approx 8-6$). This period is called the Epoch of Reionization (EoR). A brief history of these events is sketched in Fig. 1.1.

Thus, the EoR corresponds to the transition from a neutral to an ionized Universe. The main sources of the ionizing radiation are considered to be the young, star-forming galaxies. The reionization process itself has three distinct stages (Fig. 1.2). The initial stage is the "preoverlap" phase, which consists of individual galaxies ionizing their surrounding regions and forming the first HII bubbles. The first galaxies form in the most massive haloes which are located in the highest-density regions - which have a high recombination rate. Because of that, the ionization fronts propagate slowly through the high-density regions, and faster once they reach lower-density regions, thus ionizing voids faster than high-density regions. At this point, the IGM is characterized by ionized regions which are separated from neutral regions by ionization fronts [73]. First sources are highly clustered and thus the reionization proceeds fast towards the second "overlap" phase where neighbouring HII regions begin to overlap forming larger ionized regions with more sources of ionizing radiation. The final "postoverlap" phase has a slower evolution in the ionization background. This is because once most of the Universe is ionized the only still neutral regions are dense pockets of the IGM where the recombination is much higher. These regions are high-redshift analogues of Lyman Limit Systems (LLS), which absorb any ionizing photon crossing their path, slowing the growth of HII regions [73]. The reionization process does not evolve equally fast throughout the Universe. Some regions have an overabundance of sources, thus the reionization process evolves more rapidly in those regions, while some other regions have fewer sources, requiring additional photons from neighbouring regions to complete their reionization [73].

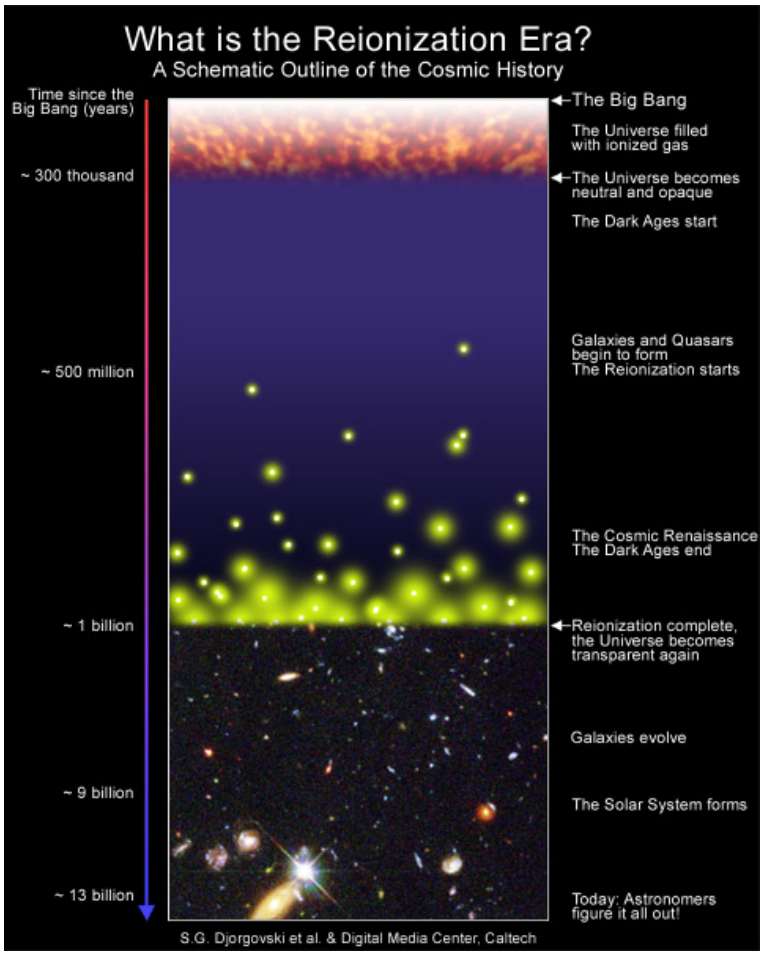

Figure 1.1: This image shows a brief history of the Universe. After the Big Bang the Universe is slowly cooling down and around 300 000 years after the Big Bang the process of recombination takes place, which results in the decoupling of matter and radiation. This relic radiation is observed as CMB. After the recombination the Universe is neutral in the period known as the Dark Ages. At the end of the Dark Ages ($\sim$ 500 million years after the Big Bang) the first stars and galaxies form, heating up the Universe and starting the Epoch of Reionization. The reionization is complete around 1 billion years after the Big Bang (redshift $z \approx 6$). This image is acquired from MIT Haystack Observatory and designed by Djorgovski et al.

Absorption spectra of high-redshift quasars suggest that reionization was completed by $z \approx 6$ [30, 7, 81]. On the other hand, measurements of the primordial CMB radiation obtained by the Planck satellite suggest that the Universe was neutral until $z = 8.8 \pm 1.1$, if instantaneous reionization is assumed [93]. No observations have yet offered much information on the progress of reionization, thus most of our information on the evolution of reionization and the properties of its sources come from theoretical calculations and simulations. However, many efforts are on the way to rectify that.

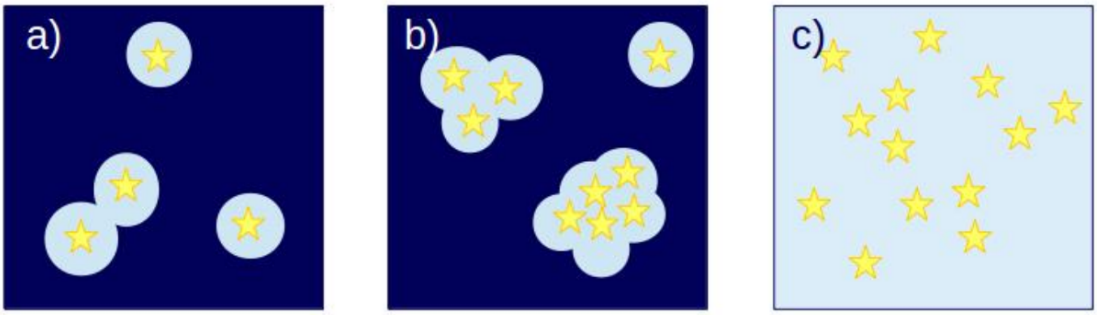

Figure 1.2: Phases or the reionization process: a) "Preoverlap" phase. This phase consists of individual galaxies ionizing their surrounding regions and forming the first HII bubbles.; b) "Overlap" phase. As the first sources are highly clustered they soon reach the "overlap" phase where their individual bubbles merge into fewer larger bubbles with more sources of ionizing photons.; c) "Postoverlap" phase. In the "postoverlap" phase most of the Universe is ionized and the only still neutral regions are dense pockets of the IGM where the recombination is much higher.

Some of the different efforts to constrain the Epoch of Reionization are with CMB measurements, quasi-stellar object (QSOs) and gamma ray burst (GRB) absorption lines, 21cm emission, and the galaxies. In this thesis, I will mostly focus on the measurements of the 21 cm emission and absorption, and the Lyman Alpha Emitters (LAEs) observations. There are significant efforts to detect reionization by mapping the 21cm line of neutral hydrogen with radio arrays such as LOFAR[1] [116], MWA[2], HERA[3] and SKA[4]. They aim to measure the intensity of the 21 cm line from intergalactic neutral hydrogen compared to CMB at different redshifts, whose topology could give us great insights into the evolution of the Epoch of Reionization, of sources of ionizing radiation, and of the sizes of ionized regions. Calculations predict that the cosmological 21cm signal from the EoR will be extremely faint, while the system noise and the foregrounds will be orders of magnitude larger (e.g. [54, 64, 46, 128, 5, 94]). Due to the low signal to noise ratio (which for LOFAR is ~ 0.2, [64]) the first observations of the 21cm signal will measure only statistical properties, such as the rms and power spectrum of the brightness temperature and their evolution with time (e.g. [18, 3, 54, 96, 47, 46]). However, cross-correlation of the 21cm signal with observations at different wavelengths such as near-infrared background radiation (NIRB; e.g. [32, 79]), kinetic Sunyaev-Zel'dovich effect (kSZ; e.g. [53, 111]), galaxies (e.g. [71, 121, 87]), CO line (e.g. [117, 70]) and CII line (e.g. [105, 127]) can provide further insight into different aspects of the EoR, such as the progress of reionization and the redshift at which the process is halfway, the evolution of the neutral hydrogen content, and the typical scale

[1]http://www.lofar.org
[2]http://web.haystack.mit.edu/arrays/MWA
[3]https://reionization.org
[4]http://www.skatelescope.org

of ionized regions at different redshifts.

When probing the Epoch of Reionization with galaxies the focus is on high-redshift, young, star-forming galaxies. Such galaxies are expected to have a strong Lyman-α emission line due to the interaction of the interstellar medium with ionizing radiation from young massive stars [88]. Depending on the detection method, such galaxies are typically referred to as Lyman-α emitters (LAEs) and Lyman break galaxies (LBGs). Star-forming galaxies that are luminous enough to be detected with existing telescopes most likely populate fairly massive dark matter haloes, with masses in excess of 10^{10} $M_\odot$ [26]. They ionize their surroundings forming large HII bubbles in which one or more star-forming galaxies reside (e.g. [26]). Lyman-α photons emitted by those galaxies can therefore propagate and redshift away from line resonance through the ionized IGM before entering the neutral IGM (e.g. [100, 51, 85, 20]). These photons are then less likely to be scattered out of the line of sight. This is why LAE luminosity functions (e.g. [45, 77, 28, 56, 55]), number density [78] and clustering [41, 124, 56] are the main methods to study the EoR with galaxies. A reduction in the number of observed sources, and thus a suppression of the luminosity function, is expected with increasing redshift, due to the larger amount of neutral gas in the IGM (e.g. [45]).

Important tools used for understanding the EoR and make plans and predictions for the upcoming observations are numerical simulations and semi-numerical models. Until recently reionization on large scales had only been studied by semi-numerical models (e.g. [60, 99]). These models are computationally much cheaper than full numerical simulations, however, they do not properly account for some important processes, such as recombination, suppression of low-mass sources, non-linear halo clustering, or radiative transfer [48], which are better captured in simulations. For accurate modeling of reionization all relevant physical processes should be included, and large scales, as well as high resolution, should be attained. Simulations of reionization were initially only a few Mpc in size (e.g. [43, 14, 42]). However, due to the advancement of technology, novel codes for cosmological N-body and hydrodynamical simulations and for radiative transfer finally enable reionization simulations with volumes larger than $\sim$ 100 Mpc [49, 82, 48], allowing for a correct abundance of rare massive haloes [2, 69, 115], while resolving also dwarf-size galaxies with masses $\sim 10^8$ $M_\odot$, which are considered to be the main sources of ionizing photons [118, 72, 98, 34]. In addition, these simulations are better suited to capture large ionized regions (which are expected to have sizes of tens of comoving Mpc towards the end of reionization; e.g. [84]) and to be compared to wide-field surveys of high redshift sources, such as the one currently conducted by LOFAR [48].

1.2 Lyman-α line from high-redshift galaxies

1.2.1 Lyman-α line

Young, star-forming galaxies produce strong Lyman-α emission making the Lyman-α line a very important probe of the high-redshift galaxies. The Lyman-α line is a hydrogen spectral

line for a $2p \rightarrow 1s$ transition and has a wavelength of 2.47×10^{15} Hz. In high-redshift galaxies the Lyman-α line is emitted when hydrogen atoms in the interstellar medium recombine after being ionized by hot, massive stars. How strong is the line luminosity of a Lyman-α emitting galaxy depends on the initial mass function (IMF, which determines what fraction of the stellar mass ends in massive, hot stars), the star formation rate $\dot{M}_\star$, the ionizing photon production rate Q_i, and the escape fraction f_{esc} which is the fraction of the ionizing photons that escape the host galaxy [73]. The intrinsic line luminosity of a galaxy is

$$L_{Ly\alpha}^{int} = \frac{2}{3} Q_i h \nu_\alpha (1 - f_{esc}) \dot{M}_\star, \tag{1.1}$$

where ν_α is the frequency of the Lyman-α line and h is the Planck constant [73]. The Lyman-α line is very optically thick in this environment, thus the Lyman-α photons scatter many times before they can escape the galaxy, and once they leave it they can eventually be scattered out of the line of sight [73]. This scattering can change the overall brightness of the line, its frequency structure, and the relation to the galaxy's continuum photons. The observed line luminosity is then

$$L_{Ly\alpha}^{obs} = \frac{2}{3} T_{Ly\alpha}^{IGM} T_{Ly\alpha}^{ISM} Q_i h \nu_\alpha (1 - f_{esc}) \dot{M}_\star, \tag{1.2}$$

where $T_{Ly\alpha}^{ISM}$ is the fraction of Lyman-α photons that are transmitted through the galaxy's interstellar medium (ISM), and $T_{Ly\alpha}^{IGM}$ is the fraction transmitted through the IGM [73].

1.2.2 Radiative transfer of Lyman-α photons

As Lyman-α photons travel through the ISM they can scatter many times (changing both their direction and frequency). Scattering cannot destroy them, but dust absorption can. Depending on the properties of the ISM, the increased path length can either increase or decrease the brightness of the Lyman-α line relative to the continuum and scattering can change the shape of the emerging Lyman-α line.

If the Lyman-α line emerges from an optically thick environment the photons will scatter many times and only in the damping wing of the line they can escape. Thus the emerging Lyman-α emission will have a double-peaked profile: photons near line center are trapped in the cloud and only when they diffuse to large positive or negative velocity are they able to escape (Fig. 1.3) [73].

If the environment from which the Lyman-α emission emerges has a velocity gradient due to either expansion (arising from winds, which are common in the star-forming galaxies), or contraction (from the infall of surrounding material), the scattered photons can acquire a redshift [73]. When the expansion velocity is much larger than the thermal velocities the photons with large positive frequency jumps cannot escape and thus only a single emission line on the red side is emerging [73]. In a contracting medium, photons typically acquire a blueshift, producing a single emission line on the blue side (Fig. 1.3) [73].

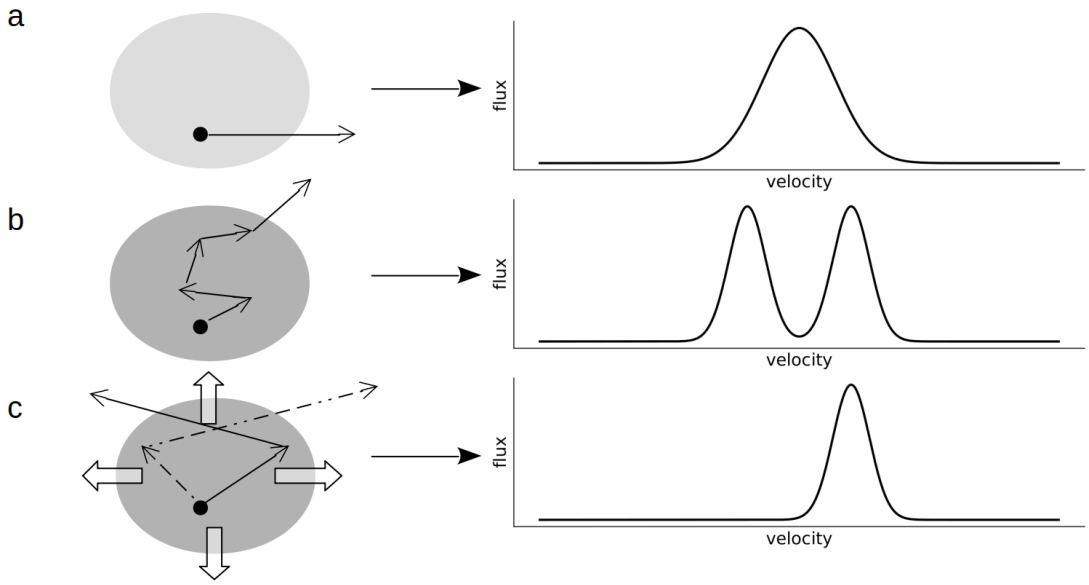

Figure 1.3: Scattering of the Lyman α line: (a) the unscattered Lyman-α line (b) Lyman-α line emerging from an optically thick environment (c) Lyman-α line emerging from an expanding optically thick environment.

Lyman-α emission emerging from a dusty environment will have many of its photons destroyed. To escape the HI cloud, the photon must first scatter far into the wings of the line and then remain in the wings as it spatially diffuses out of the system [73]. During that process, the photon will scatter many times, with a great probability that it will be absorbed by dust and destroyed in the damping wing [73]. In a uniform medium, line photons are more affected by dust than continuum photons because the many scatterings they go through are forcing them to follow a path length much longer than continuum photons, thus providing a much larger opportunity for destruction by dust [73].

Most galaxies are not just optically thick and uniformly full of dust. Instead, they can be described as a multiphase medium with dust, where multiphase means a medium in which both the HI and dust are confined to optically thick, discrete clouds, separated by a highly ionized, dust-free medium. An inhomogeneous medium will allow more transmission than a homogeneous medium with an identical column density of the neutral gas [73]. In some cases, the line photons can be less affected than continuum photons due to line photons scattering off the surface of the clouds, while the continuum photons go through them where they can encounter more dust [73]. Thus, Lyman-α photons suffer relatively few scattering events inside each cloud before spatially diffusing out. The photons can then travel a large distance before encountering another cloud, and spatial diffusion through the ionized, dust-free medium provides most impact towards escape [73].

Once the Lyman-α line finally exits the high-redshift galaxy, with its changed shape, it traverses the IGM which is itself full of neutral hydrogen at the redshifts of interest. Thus,

depending on the position of its host galaxy, the Lyman-α line is likely to be also scattered in the IGM before reaching the observer.

The full cross-section for a Lyman-α photon scattering over a single atom is

$$\sigma_\alpha(\nu) = \frac{3\lambda_\alpha^2\Lambda_\alpha^2}{8\pi}\frac{(\nu/\nu_\alpha)^4}{4\pi^2(\nu-\nu_\alpha)^2 + (\Lambda_\alpha^2/4)(\nu/\nu_\alpha)^6}, \tag{1.3}$$

where $\Lambda_\alpha = 8\pi^2e^2f_\alpha/3m_ec\lambda_\alpha^2 = 6.25\times10^8$ s^{-1} is the Lyman-α ($2p \rightarrow 1s$) decay rate, $f_\alpha = 0.4162$ is the oscillator strength, and $\nu_\alpha = c/\lambda_\alpha = 2.47\times10^{15}$ Hz is the frequency of the Lyman-α line [73]. The term in the numerator is responsible for classical Rayleigh scattering. Approximating the line as narrow

$$\sigma_\alpha(\nu) = \frac{3\Lambda_\alpha\lambda_\alpha^2}{8\pi}\delta(\nu-\nu_\alpha), \tag{1.4}$$

defining r as the photon's proper distance from the observer, and defining neutral hydrogen density as $n_{HI}(z) = x_{HI}n_H(z)$, where x_{HI} is the HI fraction, and n_H is the hydrogen number density at redshift z then

$$\begin{aligned}\tau_\alpha &= \int \sigma_\alpha(r)n_{HI}(r)dr \\ &= \frac{c}{H_0}\int\frac{da}{a}\sigma_\alpha(\nu_{obs}/a)n_{HI}(a)[\Omega_m/a^3+\Omega_\Lambda]^{-1/2} \\ &= \frac{3\Lambda_\alpha\lambda_\alpha^3}{8\pi}\frac{x_{HI}n_H(z)}{H(z)} \\ &\approx 1.6\times10^5 x_{HI}(1+\delta)\left(\frac{1+z}{4}\right)^{3/2},\end{aligned} \tag{1.5}$$

where $dr = c\ dt = c\ da/\dot{a} = c\ (da/a\ H)$ with the Hubble parameter $H = \dot{a}/a$ evaluated in the matter-dominated era, and $n_H(z) = \bar{n}_H(z)(1+\delta)$ is the mean cosmic density at redshift z [73]. The average IGM optical depth in the Lyman-α transition is referred to as the Gunn-Peterson optical depth. As the IGM optical depth can be very high even when the neutral HI fraction is small, any transmission of the Lyman-α line thus means that the IGM is highly ionized.

The fate of photons emitted within (or near) the Lyman-α line of a galaxy or quasar depends on the relative velocity and broadening of the line from bulk, thermal, or turbulent motions - which determines whether the photons pass through the Lyman-α resonance (and thus experience the full Gunn-Peterson absorption) or remain redward of the line center, experiencing less absorption [73]. It is also very relevant whether the source is embedded in completely neutral gas or in an ionized bubble, and the surrounding velocity field [73].

Once the Lyman-α line emerges from galaxies, the photons that are slightly blueward of line centre redshift into the Lyman-α resonance near their source [73]. In most models, this nearby region will already have been ionized either by the source itself or by its neighbours (if it is part of a much larger ionized bubble). However, since $\tau_\alpha > 10^5\ x_{\rm HI}$ at these

redshifts, even in a highly ionized medium the absorption can be substantial. The short mean free paths in the HII regions at high redshifts, which can be restricted either by the size of the ionized region or by LLS, will most likely prevent the gas from becoming extremely ionized [73].

The total flux in the Lyman-α line emerging from a galaxy provides a measure of its star formation rate, while its exact shape contains information on the sources powering the emission and on the gas distribution and kinematics through which the Lyman-α photons propagate [27]. Due to the complex physics behind the Lyman-α line's radiative transfer through ISM and IGM, the line is not considered a good measure of star formation rate [73]. However, its extreme brightness in many galaxies makes it a great tool for the detection of high redshift galaxies.

1.2.3 Observations of the high-redshift galaxies via Lyman-α line

High-redshift galaxies, which are very faint and thus cannot be observed in large numbers in standard ways, are mostly observed by one of two methods specifically designed for their detection.

The first method is observing the sky with several broad photometric bands to estimate the redshift of galaxies, which is based on the strong spectral break due to the absorption by intergalactic neutral hydrogen along the line of sight to the source. As the IGM is very optically thick to Lyman-α photons at the high redshifts in question, little or no flux should be detectable at frequencies higher than 1216 $(1 + z)$ Å [73]. To identify a high-redshift galaxy with this method two filters are needed - one above and one below the Lyman-α break of the galaxy defined as above. The galaxies thus observed are called Lyman-Break Galaxies. This method is widely used, however, the redshifts of the observed galaxies are not very precisely defined (often with $\Delta z \sim 1$).

The second method is using a narrow band filter to identify high-redshift galaxies whose specific emission line falls within the selected frequency band, but which are otherwise faint (or invisible) in the nearby broadband filter [73]. This method is particularly applied to the Lyman-α line, which is very prominent in high-redshift galaxies. The galaxies observed by this method are named Lyman-α Emitters (LAEs). The advantage of this method is that narrow band filters can determine quite precisely the redshift of a galaxy (e.g. for Subaru $\Delta z = 0.1$) as well as its position on the sky. One of the main challenges of this method is the contamination from low-redshift line emitters. Other strong emission lines from lower-redshift galaxies, such as Hα, [OIII], Hβ, [OII] lines, can contaminate the LAE observations. One way of eliminating these interlopers from the selection process is by contrasting the shapes of the observed lines. Lyman-α line is very asymmetric with a sharp cutoff on the blue side (due to IGM absorption) and a long tail to the red side (due to radiative transfer effects), whereas metal lines are generally very symmetric [73]. For a more robust identification, multiple lines from the same source need to be observed or a deep follow-up made to observe the continuum break. As this method defines the redshifts of the galaxies more precisely I will hereafter focus more closely on the LAEs.

Around the end of reionization the ionized bubbles are so large that the Lyman-α

emission from most galaxies can redshift away from the line resonance before reaching the end of the ionized bubble, thus we can observe most of the LAEs. However, as we go to higher redshifts, the bubbles are smaller, and thus an increasingly larger fraction of Lyman-α lines are scattered away and fewer LAEs are observed. However, there are also other reasons the LAE number density could decline with increasing redshift. The halo mass function changes rapidly with redshift during these early times of galaxy evolution, so the galaxy abundance does too [73]. Thus, the number of galaxies, including LAEs, at different redshifts can change considerably. One way to account for the halo mass function evolution is to calibrate the experiment to a broad band galaxy survey that is not subject to the same selection effects [73]. If then the number of LAEs declines fast with the increasing redshift, while the overall galaxy density observed by a broadband survey only shows a gentle decline, it would be good evidence for an IGM absorption and thus reionization.

The Lyman-α line emission from the galaxies within large ionized bubbles can remain unattenuated while the emission from the galaxies inside small bubbles will most likely be extinguished by the damping wing absorption from the IGM. This suggests that not only the mean number density of LAEs evolves throughout reionization, but their spatial distribution does too. The best way to explain the evolution of their spatial distribution is through the clustering of the galaxies. Due to the bias of the underlying dark matter field, the overdense regions tend to reside near the other overdense regions and thus other large ionized bubbles. As a consequence, we are more likely to see a lot of galaxies together than if they were randomly distributed. That is why the ionized bubbles introduce a scale-dependent bias to the correlation function of galaxies, with a break at $r \approx R_c$, where R_c is the characteristic size of the ionized bubbles [73]. Thus, the correlation function and the power spectrum are the most common measure of clustering. Another way to test for clustering is with counts in cells - dividing the survey volume into small cells and examining the distribution of the galaxy counts in the cells [73]. Clustering is also important for its effects on sample variance - a number counts of galaxies within surveys of limited volume, which is a result of the fact that a survey field can lie in a region of high galaxy density or an underdense region or a void [73]. It is important to account for it for precise estimates of luminosity functions and stellar mass functions.

In this thesis I will focus on the LAE observations to cross-correlate them with the 21cm emission maps of the intergalactic neutral hydrogen at the same redshifts. The surveys observing the LAEs, e.g. Subaru and WFIRST, will only be able to observe the brightest LAEs which are great tracers of larger ionized regions. It is expected that these LAEs should thus be anti-correlated with the 21cm emission maps on large scales since the intergalactic neutral hydrogen is generally found outside of the ionized regions. More on the information that can be extracted from such a cross-correlation can be found in Chapter 2. Once the Lyman-α line finally exits the high-redshift galaxy, with its changed shape, it traverses the IGM which is itself full of neutral hydrogen at the redshifts of interest. Thus, depending on the position of its host galaxy, the Lyman-α line is likely to be also scattered in the IGM before reaching the observer.

1.3 21cm line

1.3.1 Basic definitions

The 21cm line is a hyperfine transition line of the neutral hydrogen between the two $1s$ states (Fig. 1.4). It has a rest frequency of $\nu_0 = 1420.4057$ MHz and wavelength of $\lambda = 21.1061$ cm. How much of the neutral hydrogen is in the excited state is determined by the spin temperature, T_S, which quantifies the ratio of number densities of atoms in the hyperfine levels of the electronic ground state:

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{-E_{10}/k_B T_S} = 3\ e^{-T_\star/T_S}, \tag{1.6}$$

where n_0 and n_1 are relative number densities of the singlet and triplet states respectively, $g_0 = 1$ and $g_1 = 3$ are corresponding statistical weights, $E_{10} = 5.9 \times 10^{-6}$ eV is the energy splitting, and $T_\star \equiv E_{10}/k_B = 0.68$ K is the equivalent temperature [39]. All astrophysical applications have $T_S \gg T_\star$, thus approximately three of four atoms are in the excited states [39].

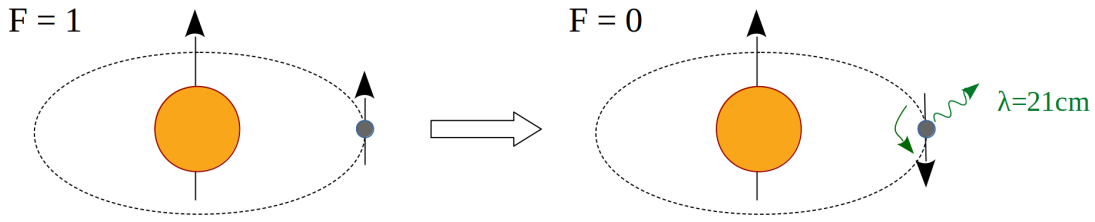

Figure 1.4: The 21cm hyperfine transition line is emitted when the electron in the $1s$ state is parallel to the nucleus, $F = 1$ state, and spin-flips to disalign with the spin of the nucleus, $F = 0$ state.

When observing the 21cm signal what is actually being observed is the differential brightness temperature of the contrast between high-redshift 21cm emission and the CMB, which is defined as:

$$\begin{aligned} \delta T_b(\nu) &= \frac{T_S - T_{CMB}(z)}{1+z}(1 - e^{-\tau_{\nu_0}}) \approx \frac{T_S - T_{CMB}(z)}{1+z}\tau_{\nu_0} \\ &\approx 9 x_{HI}(1+\delta)(1+z)^{1/2}\left[1 - \frac{T_{CMB}(z)}{T_S}\right]\left[\frac{H(z)/(1+z)}{dv_\parallel/dr_\parallel}\right]\ \text{mK}, \end{aligned} \tag{1.7}$$

where $T_{CMB}(z)$ is the CMB temperature at redshift z, τ_{ν_0} is the optical depth of a hydrogen cloud at frequency ν_0, x_{HI} is the neutral hydrogen fraction, $(1+\delta)$ is the fractional baryon overdensity, $H(z)$ is the Hubble parameter, and $dv_\parallel/dr_\parallel$ is the gradient of the proper velocity along the line of sight [39]. The observability of the 21cm signal depends on T_S,

which also determines whether the 21cm signal will be seen in emission ($\delta T_b(\nu) > 0$) or absorption ($\delta T_b(\nu) < 0$).

In absorption, the 21cm signal can also be observed as absorption features in the spectrum of a strong high-redshift radio source. Such a source, e.g. a radio-loud quasar or a radio galaxy, will have a brightness temperature as high as $T_{source} \approx 10^6 - 10^{10}$ K, which exceeds T_S and T_{CMB} by far. Thus, the flux density coming from a high redshift radio source would depend on the optical depth of the IGM along the line of sight, $S_\nu \approx S_{source} e^{-\tau_\nu}$ [39]. These high-redshift radio-loud quasars and radio galaxies would thus be great probes of small scale structure in the (partially) neutral intergalactic medium.

1.3.2 Excitation mechanisms

As the spin temperature determines whether the 21 cm signal can be observed, and if the signal will be observed in emission or absorption, it is important to understand the processes that determine the T_S itself. There are three processes that determine T_S: absorption or emission of the photons from the CMB or other background radio source, collisions with other hydrogen atoms and free electrons, and scattering of Lyman-α photons that can cause a spin-flip transition in hydrogen atoms [39, 97].

If only CMB photons are present, the spin states reach a thermal equilibrium where $T_S = T_{CMB}$ on a timescale of $\sim T_\star/(T_{CMB} A_{10}) = 3 \times 10^5 (1+z)^{-1}$ yr, which is much shorter than the age of the Universe at all redshifts after cosmological recombination [73]. When this equilibrium occurs there is no 21 cm signal. However, collisions and scattering of Lyman-α photons break this coupling. In other words, the spin temperature is determined by the equilibrium condition

$$n_1(C_{10} + P_{10} + A_{10} + B_{10} I_{CMB}) = n_0(C_{01} + P_{01} + B_{01} I_{CMB}), \tag{1.8}$$

where C_{10} and P_{10} are de-excitation rates (per atom) from collisions and UV scattering, respectively, C_{01} and P_{01} are the corresponding excitation rates, A_{10}, B_{10} and B_{01} are the appropriate Einstein coefficients, and I_{CMB} is the specific intensity of CMB photons [39]. As the Rayleigh-Jeans approximation is valid in this regime, Eq. 1.8 can be reformulated as

$$T_S^{-1} = \frac{T_{CMB}^{-1} + x_c T_{gas}^{-1} + x_\alpha T_c^{-1}}{1 + x_c + x_\alpha}, \tag{1.9}$$

where x_c and x_α are coupling coefficients for collisions and Lyman-α scattering, respectively, and T_{gas} is the gas kinetic temperature [39, 97].

Collisional excitation and de-excitation of hyperfine levels dominate in dense gas, and the gas density was high in the early Universe. Collisions occur either between two hydrogen atoms or between a hydrogen atom and a proton or an electron. The coupling coefficient for collisions with species i is

$$x_c^i \equiv \frac{C_{10}^i}{A_{10}} \frac{T_\star}{T_{CMB}} = \frac{n_i \kappa_{10}^i}{A_{10}} \frac{T_\star}{T_{CMB}}, \tag{1.10}$$

where κ_{10}^{i} is the rate coefficient for spin de-excitation in collisions with species i, with units of cm^3 s^{-1}, and species denoted with i are hydrogen atoms, free electrons, protons, and other species (helium and deuterium) [39, 97]. The total x_c is the sum over all species i [39]. Collisional coupling between neutral hydrogen atoms dominates during the Dark Ages, at $z > 30$, when the ionization fraction is still small, while free electrons become important when the Universe becomes significantly heated and ionized [39]. However, as Lyman-α photon scatters 10^5 times before redshifting out of resonance and only 10 % of collisions lead directly to de-excitation, resonant scattering of Lyman-α photons and thus the Wouthuysen-Field effect dominates the spin temperature couplings once the first galaxies form [39, 73, 97].

The Wouthuysen-Field effect is the mechanism that couples the spin temperature of the neutral hydrogen to a Lyman-α photon through the resonant scattering. The mechanism of the Wouthuysen-Field effect is shown in Fig. 1.5. If a hydrogen atom in a hyperfine singlet

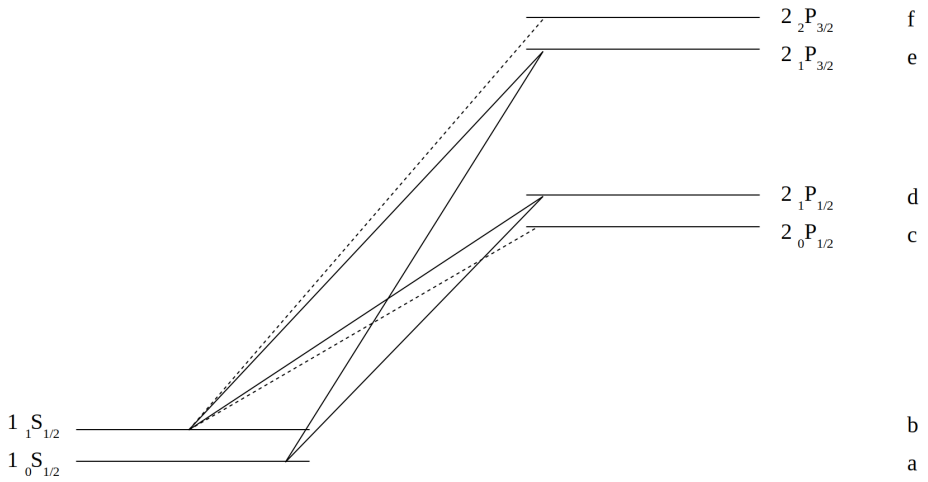

Figure 1.5: Level diagram demonstrating the Wouthuysen-Field effect due to which hydrogen atoms can change hyperfine states through the absorption and spontaneous re-emission of a Lyman-α photon. a-f symbols denote 1S and 2P hyperfine levels in order of increasing energy.

state absorbs a Lyman-α photon the atom would be excited to either of the central $2P$ states, due to the fact that the electric dipole selection rules allow total angular momentum change $\Delta F = 0, 1$, except $F = 0 \rightarrow 0$ which is prohibited [39]. The same rules will allow the $2P$ state to decay to the either of the ground states once it emits the Lyman-α photon [39, 97]. Since the atom was at the hyperfine singlet state before the absorption of the Lyman-α photon, the decay to the triplet state, ${}_1S_{1/2}$ state, signifies a change in the hyperfine state. In other words, hydrogen atoms can change their hyperfine state through the absorption and spontaneous re-emission of a Lyman-α photon [39]. The Wouthuysen-Field coupling depends on the total rate at which Lyman-α photons scatter from a hydrogen

atom

$$P_\alpha = 4\pi\chi_\alpha \int d\nu J_\nu(\nu)\phi_\alpha(\nu), \tag{1.11}$$

where $\sigma_\nu \equiv \chi_\alpha\phi_\alpha(\nu)$ is the local absorption cross-section, $\chi_\alpha \equiv (\pi e^2/m_e c)f_\alpha$, $f_\alpha = 0.4162$ is the oscillator strength of the Lyman-α transition, $\phi_\alpha(\nu)$ is the Lyman-α absorption profile, and J_ν is the angle-averaged specific intensity of the background radiation field [39].

To relate the total scattering rate of Lyman-α photons, P_α, to indirect de-excitation rate, P_{10}, 1S and 2P hyperfine levels will be labelled in order of increasing energy with a-f, and the background flux at the frequency corresponding to the $i \to j$ transition is J_{ij}. Then it follows

$$P_{01} \propto B_{ad}J_{ad}\frac{A_{db}}{A_{da}+A_{db}} + B_{ae}J_{ae}\frac{A_{eb}}{A_{ea}+A_{eb}}, \tag{1.12}$$

where A_{ij} and B_{ij} are spontaneous emission and absorption coefficients for transitions between these levels [39]. The first term is the probability for an $a \to d$ transition, $B_{ad}J_{ad}$, coupled with the probability that the subsequent decay will terminate in state b. The second term displays the same probabilities but for transitions to and from state e [39]. To relate individual A_{ij} coefficients to the total Lyman-α spontaneous emission rate, $A_\alpha = 6.25\times10^8$ Hz, which is averaged over all hyperfine sublevels, a sum rule is used which defines that the sum of decay intensities $g_i A_{ij}$ for transitions from given nFJ to all the $n'J'$ levels is proportional to $2F+1$ [6, 39]. If the permitted transitions are ordered as (bc, ad, bd, ae, be, bf), where the labels represent the initial and final states, then the relative strengths of these transitions are $(1, 1, 2, 2, 1, 5)$ [39]. Assuming that the background radiation field is constant across the individual hyperfine lines, then $P_{10} = (4/27)P_\alpha$. The coupling coefficient x_α is then

$$x_\alpha = \frac{4P_\alpha}{27A_{10}}\frac{T_\star}{T_{CMB}} = S_\alpha\frac{J_\alpha}{J_\nu^c}, \tag{1.13}$$

where S_α is a correction factor that accounts for variations in the intensity near the line center, and J_ν is evaluated at line center and set as $J_\nu^c \equiv 1.165\times10^{-10}[(1+z)/20]$ cm^2 s^{-1} Hz^{-1} sr^{-1} [39]. The coupling threshold for $x_\alpha = S_\alpha$ can be written as a number of Lyman-α photons per hydrogen atom in the Universe, $\tilde{J}_\nu^c = 0.0767[(1+z)/20]^{-2}$ [39].

Lyman-n photons can also contribute to the Wouthuysen-Field effect. A Lyman-n photon will typically scatter $N_{scatter} \sim 5$ times before initiating a decay cascade, unlike Lyman-α photon which scatters $\sim 100,000$ times before being destroyed. Consequently, coupling from the direct scattering of Lyman-n photons is suppressed by a factor $\sim 10^{-6}$ compared with Lyman-α [39]. However, Lyman-n photons can still be important because the original Lyman-n photon can be 'recycled' into Lyman-α photon, which then scatters many times through the IGM. Quantum mechanical calculations show that the higher states ($n > 3$) all have a fraction $f_{rec} \sim 1/3$ of cascades that terminate in Lyman-α photons [39, 95].

Lyman-α photons redshift out of resonance very soon after they are created and so do not contribute to the coupling far from their sources. Instead, the UV photons, which

redshift into a Lyman-series line and possibly cascade down to a Lyman-α photon, dominate the scattering [39]. Wouthuysen-Field background has very uniform intensity, however, there are some fluctuations that can be rather strong. The fluctuations depend on the sources of the photons, which are most likely star-forming galaxies. However, star-formation is not the only process that can create a Lyman-α background. Other possible sources are UV photons from quasars and collisional excitation by higher-energy X-rays [39, 73].

1.3.3 The monopole of the brightness temperature

In this subsection I will discuss the redshift evolution of the global (sky-averaged) 21cm brightness temperature. The brightness temperature can be considered as a function of four relevant quantities $T_b = T_b(T_{gas}, J_\alpha, n_H, x_i)$, where T_{gas} is the gas temperature, J_α is the specific flux of Lyman-α photons, n_H is the hydrogen density, and x_i is the volume-averaged ionized hydrogen fraction [97]. The brightness temperature is also dependent on cosmological parameters, however, this dependence will not be considered any further in this thesis. The dependence of T_b on each of these quantities can saturate, so that T_b is not sensitive any more to further variations in that quantity. This particular feature helps us separate regimes where only one of these quantities dominates the signal [97]. These separate regimes are shown in Fig. 1.6. To emphasize the different regimes I will sometimes separate them by ad-hoc redshifts which are model/source dependent and as such will not try to evaluate them.

At $200 \lesssim z \lesssim 1100$ the gas temperature is coupled to the CMB temperature, $T_{gas} = T_{CMB}$, via Compton scattering of the residual free electrons remaining after recombination [97]. The gas density is high, so the spin temperature is collisionally coupled to the gas temperature and consequently to the CMB temperature, $T_S = T_{gas} = T_{CMB}$. The global brightness temperature is thus $\bar{T}_b = 0$ and there is no detectable 21cm signal.

The period between $40 \lesssim z \lesssim 200$ is often referred to as the Dark Ages. In this period the gas cools adiabatically as $T_{gas} \propto (1+z)^2$, while CMB cools as $T_{CMB} \propto (1+z)$, which decouples the gas temperature from the CMB temperature [97]. The spin temperature remains coupled to the gas temperature, leading to $\bar{T}_b < 0$. Thus the 21cm signal is seen in absorption. The fluctuations in T_b are due to density fluctuations, which gives an opportunity to probe the cosmological initial conditions [74].

By $30 \lesssim z \lesssim 40$ the Universe expanded enough that collisional coupling becomes inefficient and radiative coupling couples the spin temperature to the CMB temperature, $T_S = T_{CMB}$ [97]. In that regime there is again no detectable 21cm signal.

Between $z_\alpha \lesssim z \lesssim 30$, where z_α is the redshift at which Lyman-α coupling is saturated, the first radiative sources start to switch on and emit Lyman-α, in addition to UV, and X-ray photons. This period and the one that follows are known as the Cosmic Dawn. The Lyman-α photons couple the spin temperature back to the gas temperature, but initially the intensity of the Lyman-α radiation is not strong enough to heat the gas above T_{CMB} [97]. Thus $\bar{T}_b < T_{CMB}$ and the 21cm signal is seen in absorption. In this regime, fluctuations in the 21cm signal are dominated by density fluctuations and variations in the

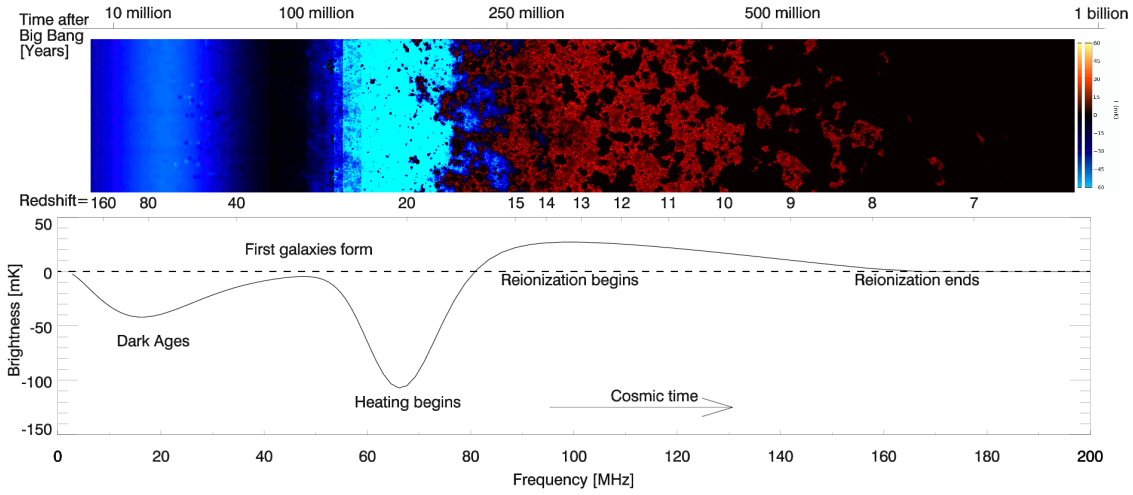

Figure 1.6: Evolution of the cosmic 21cm signal from the Dark Ages to the end of the Epoch of Reionization. Top panel: Time evolution of fluctuations in the 21cm signal. The color indicates the strength of the 21cm brightness temperature where blue colors indicate a 21cm line seen in absorption and red in emission. Black denotes where no 21cm signal can be observed. Bottom panel: Expected evolution of the global 21cm brightness temperature. There is considerable uncertainty in the exact form of this signal, due to the unknown properties of the first sources of ionizing radiation. [97]

Lyman-α flux [4, 95, 97]. By the end of this regime, when $z = z_\alpha$, the Lyman-α coupling will saturate ($x_\alpha \gg 1$) and all of the gas will be strongly coupled to Lyman-α photons [97].

In the regime between $z_h \lesssim z \lesssim z_\alpha$, where z_h is the redshift at which the gas is completely heated, the spin temperature is coupled to the gas temperature, $T_S = T_{gas}$. The Lyman-α coupling is saturated, and thus fluctuations in the Lyman-α flux do not affect the 21cm signal any longer [97]. However, heating becomes significant. In most regions $T_{gas} < T_{CMB}$ and the 21cm signal is seen in absorption, however, in some regions the gas is heated to be $T_{gas} > T_{CMB}$ and thus the 21cm signal is seen in emission. The gas is heated by Lyman-α photons and by X-rays from early accreting black holes [36] or from hot interstellar medium [86]. The fluctuations in the 21cm signal are due to gas temperature fluctuations. By redshift z_h the gas is heated everywhere to $\bar{T}_{gas} = T_{CMB}$. In this regime both astrophysical and cosmological processes that heat IGM can be probed.

The regime between $z_r \lesssim z \lesssim z_h$, where z_r is the redshift at which the Universe is completely reionized, is the Epoch of Reionization. Heating continues and brings the gas to $T_{gas} \gg T_{CMB}$. The UV radiation field ionizes the neutral hydrogen around the first sources of ionizing radiation forming the first bubbles [63]. At the beginning of the Epoch of Reionization the fluctuations in the 21cm signal are due to density fluctuations, fluctuations in ionization and the gas temperature. However, in advanced stages fluctuations in the 21cm signal are dominated by the fluctuations in the ionization field [40]. The tomography of this process is sensitive to the nature and clustering of the major sources of ionizing

radiation [82]. At the end of this epoch the 21cm signal decreases and approaches zero. During the Epoch of Reionization the nature and clustering of the sources of ionizing radiation, feedback on the IGM, and the evolution of the reionization process can be probed [63].

At $z \lesssim z_r$, hydrogen in the IGM is mostly ionized and the dominant source of the remaining 21cm signal is the galaxies [63, 97].

Most of the above regimes are not sharply defined and there could be significant overlap between them. The redshifts at which they occur are also model dependent and currently unconstrained by observations.

1.3.4 The fluctuations in the brightness temperature

The monopole of the 21cm brightness temperature contains a wealth of information, however, each component which contributes to the brightness temperature (Eq. 1.7) can fluctuate significantly [39]. The density field fluctuates due to the evolving cosmic web, and the ionization field is evolving rapidly during the Epoch of Reionization due to rapidly growing HII bubbles around galaxies which are the main sources of ionizing radiation. And Lyman-α background and X-ray heating are both strongest around galaxies. Due to the fluctuations in these components contributing to the brightness temperature, the brightness temperature will itself be highly inhomogeneous when observed over large angular scales [97]. Measuring these fluctuations in the 21 cm signal is one of the goals of the upcoming observational efforts. The most common tool to analyze these fluctuations is a power spectrum.

If $\delta_{21}(\mathbf{x}) \equiv [T_b(\mathbf{x}) - \bar{T}_b]/\bar{T}_b$ is the fractional perturbation of the brightness temperature and $\tilde{\delta}_{21}(\mathbf{k})$ is its Fourier transform, then the power spectrum of the fractional perturbations in the 21cm brightness temperature is defined as

$$\left\langle \tilde{\delta}_{21}(\mathbf{k_1})\tilde{\delta}_{21}(\mathbf{k_2}) \right\rangle \equiv (2\pi)^3 \delta_D(\mathbf{k_1} - \mathbf{k_2}) P_{21}(\mathbf{k_1}), \tag{1.14}$$

where $\delta_D(x)$ is the Dirac delta function, and the angular brackets denote an ensemble average [39].

As previously shown in Eq. 1.7 and Eq. 1.9 the brightness temperature depends on many parameters. Expansion of these equations to the linear order in each of the perturbations gives

$$\delta_{21} = \beta\delta_b + \beta_x\delta_x + \beta_\alpha\delta_\alpha + \beta_T\delta_T - \delta_{\partial v}, \tag{1.15}$$

where δ_i describes the fractional variation in a particular quantity i, where i stands for the baryonic density as b, the neutral fraction as x, the Lyman-α coupling coefficient as α, gas temperature as T, and the line-of-sight peculiar velocity gradient as ∂v [39]. The expansion coefficients β_i are

$$\beta = 1 + \frac{x_c}{x_{tot}(1 + x_{tot})}, \tag{1.16}$$

$$\beta_x = 1 + \frac{x_c^{HH} - x_c^{eH}}{x_{tot}(1 + x_{tot})}, \tag{1.17}$$

$$\beta_\alpha = \frac{x_\alpha}{x_{tot}(1+x_{tot})}, \tag{1.18}$$

$$\beta_T = \frac{T_{CMB}}{T_{gas}-T_{CMB}} + \frac{1}{x_{tot}(1+x_{tot})}\left(x_c^{eH}\frac{d\ln\kappa_{10}^{eH}}{d\ln T_{gas}} + x_c^{HH}\frac{d\ln\kappa_{10}^{HH}}{d\ln T_{gas}}\right), \tag{1.19}$$

where $x_{tot} \equiv x_c + x_\alpha$ and the collisional term is split into the two dominant components, H-e$^-$ and H-H, where necessary [39]. The coefficient β has two terms, the first describing the matter content and the second the collisional coupling efficiency in dense gas. In the equation for β_x the first term describes direct fluctuations in the ionized fraction, and the second the effect of the increased electron density on x_c. β_α stands for the fractional contribution of the Wouthuysen-Field effect to the coupling. In the equation for β_T the first term indicates the speed at which the spin temperature responds to the fluctuations in T_{gas}, while the others include the explicit temperature dependence on the collision rates [39].

1.3.5 Observations

There are two main approaches to observing the 21cm signal from the Epoch of Reionization and the Cosmic Down. The first approach is observing the global, sky-averaged 21cm signal throughout the different redshifts, where observations would reveal the evolution of the averaged 21cm signal, and indicate when the signal is seen in emission or absorption. The other approach is to observe the fluctuations of the averaged 21cm signal at different redshifts, to study their evolution and origin. However, these observations would not be able to measure the global, sky-averaged value of the 21cm signal.

Observations of the global, sky-averaged 21cm signal are done with single dish experiments, which can easily reach the required mK sensitivity. The main problem for these experiments is the foreground brightness. Galactic synchrotron foregrounds have a sky brightness in the range $200 - 10^4$ K (depending on frequency), which is orders of magnitude larger than the expected 21cm signal from the Epoch of Reionization [39]. Detecting the high-redshift 21cm signal will be extremely challenging, requiring very precise calibration over a wide frequency range and sharp localized features in the brightness temperature that can be distinguished from the smoother foreground. Such sharp localized features, while common in the observations of the fluctuations in the 21cm signal where the ionization front keeps a sharp boundary between areas with/without the 21cm signal, are not common in these observations. Thus in these single-dish observations of the sky-averaged signal, it will be much harder to distinguish between smooth foregrounds and the 21cm signal [8].

Unlike the observations aiming for the global sky-averaged signal, interferometric observations of the fluctuations in the 21cm signal should have less of a problem with distinguishing between the signal and the smooth foregrounds. However, the first generation instruments for interferometric observations of the 21cm signal from EoR and Cosmic Dawn, e.g. LOFAR and MWA, will have strong instrumental noise, which will make the imaging

impossible. Thus, statistical methods for extracting information from low signal-to-noise maps, e.g. power spectra, will be necessary.

This instrumental noise is due to the system temperature which includes contributions from the telescope, the receiver system and the sky. Noise fluctuations ΔT^N depend on bandwidth and integration time t_{int} according to the radiometer equation

$$\Delta T^N = \kappa_c \frac{T_{sys}}{\sqrt{\Delta\nu t_{int}}} \approx \frac{T_{sys}}{\sqrt{\Delta\nu t_{int}}}, \tag{1.20}$$

where $\kappa_c \leq 1$ is an efficiency factor accounting for details of the signal detection [39]. The noise level for unresolved sources is

$$\sigma_S = \frac{T_{sys}/K_\alpha}{\sqrt{\Delta\nu t_{int}}}, \tag{1.21}$$

where $K_\alpha \equiv T_\alpha/S_\nu = A_e/2k_B$ is the antenna sensitivity factor with S_ν being the source flux density and A_e the telescope collecting area [39].

The noise level decreases with increasing telescope collecting area. The total collecting area can be distributed over a larger physical area to achieve a better angular resolution, $\theta_D \approx \lambda/D_{max}$, where λ is the observed wavelength and D_{max} is the maximum separation between antennas [39]. Then, the equivalent brightness temperature uncertainty is

$$\Delta T^N = \frac{\sigma_S c^2}{2k_B\nu^2\Omega_B} \equiv \frac{T_{sys}}{n_f\sqrt{\Delta\nu t_{int}}}, \tag{1.22}$$

where $n_f \equiv A_{tot}/D_{max}^2$ is the array filling factor, and $\Omega_B \approx \theta_D^2$ is the solid angle subtended by the telescope beam [39]. The integration time needed to detect a particular surface brightness grows as $t_{int} \propto D_{max}^4$ if the total collecting area is spread over larger areas to achieve better angular resolution [39].

The sky temperature is due to the foregrounds. There are three main foreground sources that will make the detection of the 21cm signal from EoR very difficult. One of them is terrestrial radio interference. The low-frequency band for observations at redshifts $z = 6-30$ corresponds to a frequency range $\sim 45-200$ MHz. This band is heavily used by humans for analog TV stations, FM radio band, and satellite and aircraft communication channels. The resulting terrestrial radio interference is around 10 orders of magnitude brighter than the 21cm background signal [39]. This is the reason why most radio observatories are placed in isolated locations. However, this interference is usually narrow band, so it is possible to measure the cosmological signal in the gaps between contaminated channels [39].

Another foreground source is the ionosphere. The ionosphere is refractive at low frequencies and at the lowest ones, corresponding to $z > 50$, it becomes opaque [73]. The refraction phenomena cause sources to jitter across the sky as the patches of ionosphere move across the telescope beam with the timescale of several seconds. Although very complicated, it can be corrected by calibrating to the locations of a set of point sources distributed across the field of view [73]. The ionosphere is more active during the day and

during times of high solar activity. Combined with the large brightness of the sun at these frequencies it restricts the high-redshift 21cm observations to the nighttime [39].

The main foreground sources are extraterrestrial in nature. Nearly all non-thermal sources are bright in this low-frequency band [39]. However, the strongest of them is the synchrotron radiation from the Milky Way, although sources like active galactic nuclei, galaxy clusters, and radio-galaxies also contribute [73]. The high-latitude, "quiet" portions of the sky have a brightness temperature [39]

$$T_{sky} \equiv 180 \left(\frac{\nu}{180 \text{ MHz}}\right)^{-2.6} \text{ K}. \tag{1.23}$$

This "noise" is $\sim 10^4$ times stronger than the cosmic signal that is being observed [73]. Thus, the 21cm mapping will require large integration times and large collecting area to overcome this noise.

Implementations of the LOFAR and SKA noise to the various observations can be found in the subsequent chapters. In Chapter 2 I show the construction of the LOFAR and SKA observations for studies of the cross-correlations with LAEs. In Chapter 3 I show construction of SKA noise for implementations to 1D LOS power spectra observations and different configurations of 2D observations. In Chapter 4 I show implementations of LOFAR and SKA noise to the 21cm forest observations. The configuration for SKA used in this thesis is the configuration for the SKA1-LOW observations, first stage of the SKA experiment for the low frequency observations.

1.3.6 Cross-correlations with other surveys

As shown in the previous section, 21cm observations will have significant noise and foreground problems. One way of dealing with possible foreground leakage is to cross-correlate the 21cm observations with observations of some other signal from the same redshift that has different foregrounds since that would remove any foreground contamination that is not shared by both signals [73]. Besides solving leakage problems, such a cross-correlation could also offer various new insights [73]. However, it is important to keep in mind that these signals that could be cross-correlated with the 21cm signal all depend on the radiation sources, thus making these cross-correlations only possible during the Cosmic Dawn and the Epoch of Reionization; at any earlier redshift, but after the surface of last scattering, the 21cm signal is the only possible probe [39]. The 21cm signal could be cross-correlated with signals such as near-infrared background radiation (e.g. [32, 79]), kinetic Sunyaev-Zel'dovich effect (kSZ; e.g. [53, 111]), galaxies (e.g. [71, 87, 122, 119]), CO line (e.g. [117, 70]) and CII line (e.g. [105, 127]). These cross-correlations would provide further insight into different aspects of the Epoch of Reionization, such as the progress of reionization and the redshift at which the process is halfway, the evolution of the neutral hydrogen content, and the typical scale of ionized regions at different redshifts. Cross-correlations between 21cm signal and galaxies are discussed in more details in Chapter 2.

1.3.7 21 cm forest

Most observations of the 21cm line focus on mapping the 21 cm line over large scales. Unlike these observations, the 21cm forest observations would focus on small-scales in the line-of-sight to some background radio-loud source. Neutral hydrogen that is along the line-of-sight to the source would then absorb the photons that redshift into the 21cm transition, which would then create a 'forest' of absorption features in the spectrum. In these spectra we could recognize absorption features due to the diffuse IGM, filaments in the cosmic web, minihalos and ionized bubbles [39]. Since a potential bright radio source should have a high signal-to-noise spectrum, the 21cm forest observation should be, in principle, simpler than tomography [39].

However, there are a few problems that make the 21cm forest difficult to detect. As the Universe reionizes, the absorption features rapidly weaken [39]. During the reionization, the lines of sight with strong absorption features will probably be rare [39]. The filaments of the large scale structure should also become weaker because of the increase in the spin temperature and once the IGM becomes even moderately heated the cosmic filaments spikes become rare and only virialized objects have measurable absorption [39].

At the same time, the radio-loud sources should be much more common during the Epoch of Reionization than before it. As redshift increases, the radio sources should be much rarer and dimmer. The most obvious background sources for the 21cm forest would be radio-loud quasars. If we assume that their abundance declines along with their bright optical counterparts, there are expected to be 1000 radio-loud quasars at $z > 8$, and only 10 at $z > 12$ [10]. However, all the high-redshift quasars detected so far, and in particular those at the highest redshifts, are radio quiet. Other possible sources are GRBs and hypernovae. They should occur at redshifts before reionization, however, their afterglow models predict that only the most energetic events achieve the required flux densities [39]. However, GRBs of Pop-III stars could, in theory, achieve high enough flux densities at redshifts well before reionization [16, 114].

This trade-off between the decreasing trend in optical depth with progress of reionization and the decreasing radio source flux density with higher redshifts can be quantified by producing a ratio between the minimum source flux density, S_{min}, and the optical depth, τ, of the cloud of gas that is being observed through it with a signal-to-noise, S/N:

$$S_{min} = 16 \text{ mJy} \left(\frac{S/N}{5} \frac{0.01}{\tau} \frac{2500 \text{ m}^2/\text{K}}{A_{eff}/T_{sys}} \right) \left(\frac{1 \text{ kHz}}{\Delta\nu_{ch}} \frac{1 \text{ week}}{t_{int}} \right)^{1/2}, \tag{1.24}$$

where $\Delta\nu_{ch}$ is the bandwidth of each channel where it is assumed that the bandwidth is smaller than the absorption feature of interest, and t_{int} is the total integration time [39]. Thus, observing any of the narrow spectral features would require a bright radio source and long integration times on an instrument with requirements similar to SKA. The 21cm forest is discussed in more details in Chapter 4.

Chapter 2

21cm-galaxy cross-correlations

The detection of the 21cm signal from EoR is a complicated process where system noise and foregrounds will be orders of magnitude larger than the signal [54, 64, 128]. Thus, another diagnostic is needed to confirm that the detected signal is indeed from the EoR. A promising diagnostic is a 21cm-galaxy cross-correlation [37, 123], hence a cross-correlation between the sources of ionizing radiation and the inter-galactic gas that is being ionized.

Besides being a great diagnostic tool 21cm-galaxy cross-correlation also gives a wealth of information about reionization process itself. It is sensitive to the typical size of the ionized regions at the particular redshift, the ionization fraction, the clumpiness of the IGM, and the sources of the reionization [71, 106]. It can also distinguish between 'inside out' and 'outside in' reionization scenarios [121].

[71] proposed for the first time to use the shape and normalization of the 21cm-galaxy cross-power spectrum and its evolution with redshift to gain insight on the abundance of neutral gas in the IGM. [121], instead, made the first predictions for the observability of the cross-power spectrum with LOFAR, suggesting that for these studies LAEs are better suited than LBGs.

In this chapter I will continue these efforts and present predictions for the detection of the 21cm-galaxy cross-correlations with LOFAR and Subaru which both agreed to observe the ELAIS-N1 field at $z = 6.6$. I will also show predictions for observations of the 21cm-galaxy cross-correlations with the next generation instrument SKA and future galaxy surveys. The theoretical 3D 21cm-galaxy cross-power spectra are computed from full radiative transfer + N body simulations by [48], which have been post-processed with a Lyα radiative transfer code ([65, 56]) to accurately model the observed properties of LAEs. The mock 21cm and galaxy observed maps obtained by adding the LOFAR and HSC characteristics, and characteristics of SKA and future galaxy surveys, have been used to compute the 2D 21cm-LAE cross-power spectra and cross-correlation functions. The present investigation improves on previous efforts in terms of theoretical modeling (by including e.g. an accurate treatment of the radiative transfer of both line and continuum photons to model the reionization process and the properties of the LAEs), as well as mock observations (by targeting the upcoming observation of the ELAIS-N1 field by the LOFAR and Subaru telescopes, and future observations with SKA and future galaxy surveys).

2.1 Simulations

To compute the 21cm-galaxy cross-correlations I have used a full radiative transfer + N-body simulation of reionization [48] in a box of comoving length $425h^{-1}$ Mpc (corresponding to ~ 4 deg at z=7) with 165 billion particles distributed on a grid of $10,976^3$ cells ($3.9\,h^{-1}$kpc gravity force resolution) and a radiative transfer grid of 504^3 cells. The N-body simulation has been run from redshift $z = 300$ to $z = 2.6$, with initial conditions generated using the Zel'dovich approximation and a power spectrum of the linear fluctuations given by the CAMB code [68]. This simulation was then used as an input to the radiative transfer code C^2-RAY [83] to follow the reionization history of the IGM. More specifically, the halo catalogues were used to construct the sources of ionizing radiation as in [50]. As the minimum resolved halo mass is $M_{h,min} = 10^9$ M$_\odot$, haloes with masses of $10^8 - 10^9$ M$_\odot$ were modeled as a sub-grid population [48, 1]. All haloes were assigned an ionizing photon emission rate per unit time, $\dot{N}_\gamma$, proportional to the halo mass, M$_h$:

$$\dot{N}_\gamma = \frac{g_\gamma M_h \Omega_b}{\Omega_0 m_p} \left(\frac{10 \text{ Myr}}{\Delta t} \right), \tag{2.1}$$

where m_p is the proton mass, Ω_b and Ω_0 have their usual cosmological meaning, $\Delta t = 11.46$ Myr is the time between two snapshots of the N-body simulation, and g_γ is a source efficiency coefficient that incorporates the star formation efficiency, the total photon production per stellar baryon and the ionizing photon escape fraction [49, 55]. Haloes with masses down to 10^9 M$_\odot$ were assigned a source efficiency of $g_\gamma = 1.7$. Smaller sources with masses down to 10^8 M$_\odot$ were assigned $g_\gamma = 7.1$, to account for a lower metallicity and a more top-heavy initial mass function, but they were assumed to be suppressed within ionized regions (for ionization fraction higher than 10 per cent; [48]). The radiation emitted by the sources is propagated through the gridded density field, and the distribution of neutral hydrogen is obtained at various redshifts. This is used to calculate the associated differential brightness temperature according to the usual formalism (e.g. [33, 76, 39]):

$$\delta T_b = 28.5 \text{ mK } (1+\delta) x_{HI} \left(\frac{\Omega_b}{0.042} \frac{h}{0.73} \right) \times \left[\left(\frac{1+z}{10} \right) \left(\frac{0.24}{\Omega_m} \right) \right]^{1/2}, \tag{2.2}$$

where $x_{HI}(1+\delta) = n_{HI}/\langle n_H \rangle$ is the mean density of neutral hydrogen in units of the mean density of hydrogen at redshift z.

For our purposes, I used boxes from the simulation at $z = 6.68$, 7.06, 7.3, and 8.06, corresponding to volume (mass) averaged ionized fractions $\langle x \rangle = 0.93$ (0.95), 0.65 (0.73) and 0.48 (0.58), 0.21 (0.30), respectively. These particular boxes were chosen because Subaru has two narrow-band filters observing at redshifts 6.6 and 7.3, while 7.06 is an intermediate value, and WFIRST will be observing at redshift range $7.5 \leq z \leq 8.5$ for which our closest simulation box is at $z = 8.06$.

The same simulations were processed with a Lyα radiative transfer code to model high-z LAEs and study their observability. Motivated by detailed radiative transfer calculations by [66], the Lyα line was modeled as a double peaked profile with little emission at the line

centre, and a width that depends on the halo mass. Intrinsic luminosities were calibrated against observations, with a model where the Lyα luminosities of haloes of a given mass follow a log-normal distribution with a mean that is proportional to the halo mass. After assigning an intrinsic Lyα spectrum to the dark matter haloes in the N-body simulations, the observed luminosities are calculated including the attenuation from the IGM along a large number of lines of sight from each of the haloes (for more details I refer the reader to the original papers [56, 55]). From the same work I extracted the Lyα intrinsic and transmitted luminosities, which I use to produce Subaru and WFIRST mock observations.

In the computation of the 3D cross-power spectra I merged bins to obtain $\Delta k > 0.02\ h$ Mpc^{-1}, which corresponds to the smallest mode resolved by a FoV of 16 deg^2, i.e. equivalent to our simulations. I also made sure to avoid correlations in power due to the window function[1] by using a binning with $\Delta \log k = 0.02$. As the FoV used to compute the 2D cross-power spectra is smaller (i.e. 7 deg^2 and 1.7 deg^2 at z =6.6 and 7.3, respectively; see section 3.2), I used $\Delta \log k = 0.03$ (0.05) and $\Delta k > 0.04$ (0.07) h Mpc $^{-1}$ for $z = 6.6$ (7.3).

The following set of cosmological parameters was used: $\Omega_\Lambda = 0.73, \Omega_m = 0.27, \Omega_b = 0.044, h = 0.7$ and $\sigma_8 = 0.8, n_s = 0.96$, consistent with WMAP 5-year data [61].

2.2 21cm-LAE cross-correlations observable with LOFAR and Subaru

2.2.1 Cross-power spectrum

In this section I present my calculations of the theoretical and observational cross-power spectra.

At each redshift, the 21cm-galaxy cross-power spectrum at wave number $k = |k|$, $\Delta^2_{21,gal}(k)$, can be decomposed into three contributing terms (e.g. [71]):

$$\Delta^2_{21,gal}(k) = \tilde{\Delta}^2_{21,gal}(k)/\delta T_{b0} = \langle x_{HI} \rangle [\Delta^2_{x_{HI},gal}(k) + \Delta^2_{\rho,gal}(k) + \Delta^2_{x_{HI}\rho,gal}(k)], \tag{2.3}$$

where $\Delta^2_{x_{HI},gal}$, $\Delta^2_{\rho,gal}$ and $\Delta^2_{x_{HI}\rho,gal}$ are the neutral fraction-galaxy, density-galaxy and neutral density-galaxy cross-power spectra, respectively. $\tilde{\Delta}^2_{21,gal}$ is the 21cm-galaxy cross-power spectrum unnormalized by δT_{b0}, δT_{b0} is the 21cm brightness temperature relative to the CMB for a fully neutral gas element at the mean cosmic density, and $\langle x_{HI} \rangle$ is the volume-averaged neutral hydrogen fraction. $\Delta^2_{a,b}$ is the dimensionless cross-power spectrum

[1]The sphere used to compute a spherically averaged $P(k)$ in a simulation of comoving length 425 h^{-1}Mpc, must be equivalent in volume and thus have a radius $R = 264\ h^{-1}$Mpc comoving. A window function for a spherical tophat has its first zero at $dk \cdot R \sim 4.5$, so that k-values spaced by less than $4.5/R = 0.02\ h$ Mpc^{-1} will be correlated [31, 37, 71, 121]. Similarly, the circle used to compute a circularly averaged $P(k)$ must be equivalent in area and thus have a radius $R = 240\ h^{-1}$Mpc comoving. A window function for a circular tophat has its first zero at $dk \cdot R \sim 3.8$, so that k-values spaced by less than $3.8/R = 0.02\ h$ Mpc^{-1} will be correlated.

of two random fields, a and b, and it is equal to $\Delta^2_{a,b}(k) = k^3 P_{a,b}(k)/2\pi^2$ for the 3D cross-power spectrum, and $\Delta^2_{a,b}(k) = k^2 P_{a,b}(k)/\pi$ for the 2D power spectrum. $P_{a,b}$ represents the dimensional cross-power spectrum between fields a and b. The latter are represented in terms of their fractional fluctuations at a location r, i.e. $\delta_a(r) = (a(r) - \langle a \rangle)/\langle a \rangle$, and similarly for b[2]. A more detailed discussion of the various terms can be found in [71].

Theoretical 21cm-galaxy cross-power spectrum

To understand the 21cm-galaxy cross-power spectrum first is shown the theoretical spherically averaged 3D 21cm-galaxy cross-power spectrum at $z = 7.3$, together with its contributing terms (Fig. 2.1, top panel) and the corresponding cross-correlation coefficient (Fig. 2.1, bottom panel), defined as $r_{21,gal}(k) = P_{21,gal}(k)/[P_{21}(k)P_{gal}(k)]^{1/2}$. This corresponds to the ideal case in which all galaxies residing in haloes with masses $\mathrm{M}_h > 10^{10} M_\odot$ could be observed. From the behaviour of $\Delta^2_{\rho,gal}$ it is clear that, as expected, the galaxies are strongly correlated with the density field on small scales, because galaxy formation is biased toward high density regions, while the correlation decreases as we move towards larger scales, but always remains positive. The neutral hydrogen-galaxy cross-power spectrum, $\Delta^2_{x_{HI},gal}$, instead, is negative on large scales where there is a paucity of galaxies but most of the HI resides. A turn around is observed in correspondence of the typical scale of the HII regions, and then the correlation drops off since the hydrogen inside such regions is completely ionized independently from the number of sources. $\Delta^2_{x_{HI}\rho,gal}$ is positive on the largest scales and becomes negative towards smaller scales, where it cancels out with $\Delta^2_{\rho,gal}$. The final 21cm-galaxy cross-power spectrum thus follows the shape of $\Delta^2_{x_{HI},gal}$ on small scales, and that of $\Delta^2_{x_{HI},gal}$ and $\Delta^2_{\rho,gal}$ on large scales. The cross-correlation coefficient (bottom panel of Fig. 2.1) shows more clearly that the 21cm signal and the high-z galaxies are anti-correlated on large scales, and become uncorrelated on scales smaller than the typical size of the ionized regions.

Similar conclusions were drawn by [71] and [121], although our results are closer to those of [71] because of the lower resolution employed in the simulations by [121], which set a limit for ionizing photon production to halos with masses of $M_{h,min} = 10^{10}$ $\mathrm{M}_\odot$, rather than $M_{h,min} = 10^8$ $\mathrm{M}_\odot$ employed here.

Fig. 2.2 shows $\Delta^2_{21,gal}$ and $r_{21,gal}$ for the chosen redshifts. The calculation was done using galaxies inside haloes with $\mathrm{M}_h > 10^{10}$ $\mathrm{M}_\odot$, i.e. 3 million galaxies at $z = 6.68$, 2.3 million at $z = 7.06$, 1.9 million at $z = 7.3$, and 1 million at $z = 8.06$. I show that the amplitude of the power spectrum decreases with decreasing redshift, while the turnover point shifts towards larger scales. This indicates that, as reionization proceeds, the anti-correlation decreases because of the paucity of neutral hydrogen, and the ionized bubbles

[2]Note that I evaluate the theoretical cross-power spectrum with $\langle a \rangle = (\sum_{i=1}^{N} a_i)/N$, where N is the number of pixels in the portion of the simulation used. All the quantities are calculated like this, with the exception of the galaxy field in mock observations, which is instead calculated using $\langle N_{gal} \rangle = N_{gal}/V$, where N_{gal} is the number of galaxies in the mock observation, and V is the volume of the survey. This was done for an easier comparison with the shot noise power spectrum $P_{shot}(k) = 1/n_{gal}$, where n_{gal} is the average number of galaxies in the survey volume.

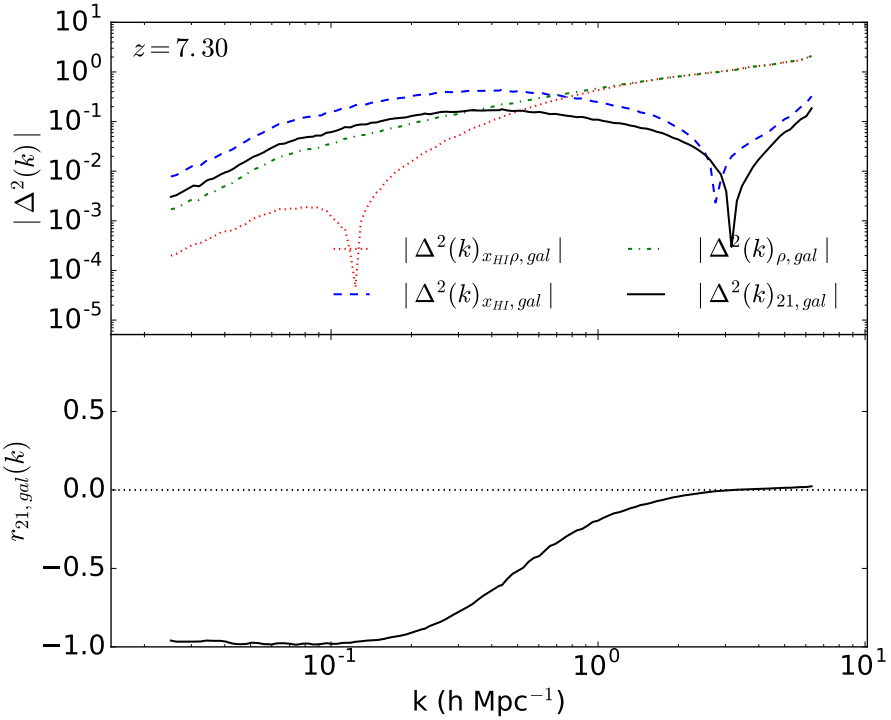

Figure 2.1: Top panel: spherically averaged 3D 21cm-galaxy cross-power spectrum, $\Delta^2_{21,gal}$ (black solid line) at $z = 7.3$, together with its contributing terms, i.e. the neutral hydrogen-galaxy cross-power spectrum, $\Delta^2_{x_{HI},gal}$ (blue dashed), the density-galaxy cross-power spectrum, $\Delta^2_{\rho,gal}$ (green dashed-dotted) and the neutral density-galaxy cross-power spectrum, $\Delta^2_{x_{HI}\rho,gal}$ (red dotted). Bottom panel: 21cm-galaxy cross-correlation coefficient, $r_{21,gal}$ (black solid), and zero-correlation coefficient (black dotted).

grow in size. This is more clearly seen in the behaviour of the cross-correlation coefficients, which shift towards smaller k with decreasing redshift.

Observed 21cm-LAE cross-power spectra

In this section I will show my predictions for the 2D 21cm-LAE cross-power spectrum as it would be observed with LOFAR and HSC. To do that, I added projection effects and constrained the galaxy number density to match HSC expectations, and I added noise to the 21cm field to simulate LOFAR observations.

HSC will probe the reionization epoch with the Ultra-deep and Deep layers of the HSC Survey. Observations are made with narrow-band filters ($\Delta z = 0.1$, equivalent to approximately one tenth of our simulation length $\approx 42\ h^{-1}$Mpc), so that the LAEs redshift will be tightly constrained. Because the LAEs detected with a particular filter will be observed as if they were lying on a single plane, the observed 21cm-LAE cross-power spectrum will be a circularly averaged 2D cross-power spectrum. HSC will observe 4 fields of 7 deg^2 at redshift $z = 6.6$ as part of a Deep layer, and 4 fields of 1.7 deg^2 (two at $z = 6.6$

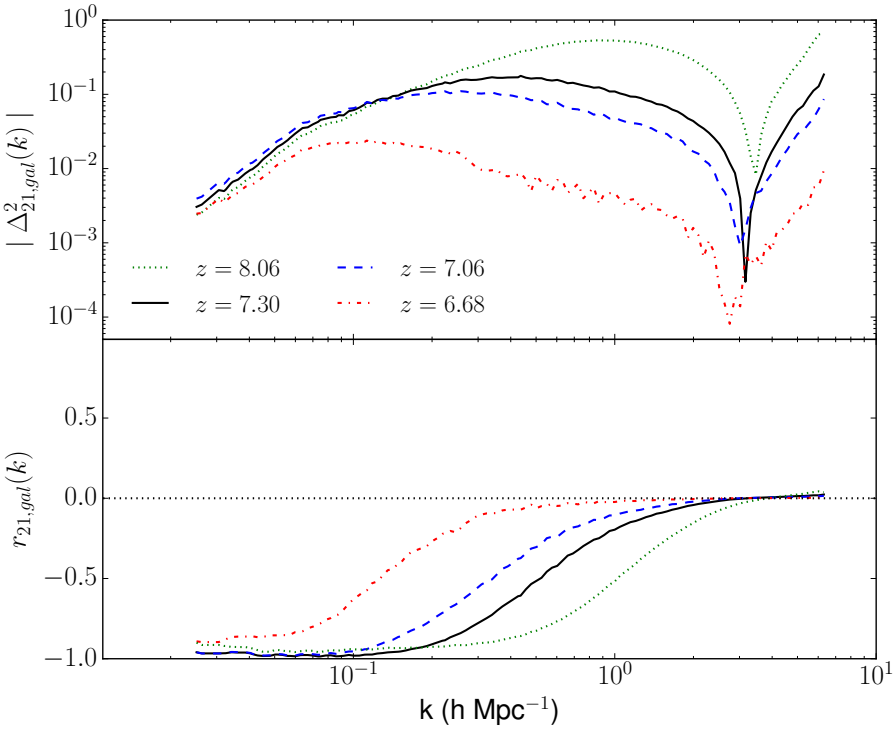

Figure 2.2: Top panel: spherically averaged 3D 21cm-galaxy cross-power spectrum, $\Delta^2_{21,gal}$, at $z = 8.06$ (green dotted line), at $z = 7.3$ (black solid), 7.06 (blue dashed), and 6.68 (red dashed-dotted). Bottom panel: 21cm-galaxy cross-correlation coefficient, $r_{21,gal}$, corresponding to $\Delta^2_{21,gal}$.

and two at $z = 7.3$) as part of an Ultra-deep layer (Ouchi 2012, private communication). One of the fields in the Deep layer is ELAIS-N1, which will also be observed with LOFAR [52].

I reduced the box dimension to match the HSC's field of view (7 deg^2 at $z = 6.6$ and 1.7 deg^2 at $z = 7.3$) by removing external cells[3]. I then divided the simulation boxes of brightness temperature and galaxies into 10 sub-boxes of 50 slices each, corresponding to a $\Delta z = 0.1$. Each sub-box obtained from the galaxy simulation is collapsed onto a single plane to mimic the fact that HSC observations will provide a 2D map of galaxies. This map is then correlated with each of the 50 slices of the corresponding brightness temperature sub-box to obtain 50 2D 21cm-galaxy cross-power spectra, which are then averaged to mimic the result of observations of a single FoV. From the 10 sub-boxes I then obtain 10 2D 21cm-galaxy cross-power spectra, which can again be averaged so that our results are not sample dependent.

Fig. 2.3 shows final, unnormalized by δT_{b0}, 2D 21cm-galaxy cross-power spectra before

[3]The choice of removing external cells is arbitrary and I have checked that it does not affect the final results.

including the noise and the constraints on the galaxy number density[4]. Even in 2D the cross-power spectra still retain much of their shape, although some features are lost due to projection effects and reduction in the field of view, e.g. the turnover point is not clear anymore. Projection effects also induce a reduction in the value of the anti-correlation, clearly observed in the cross-correlation coefficient, which drops from $r_{21,gal} \approx -1$ to $r_{21,gal} \approx -0.7$.

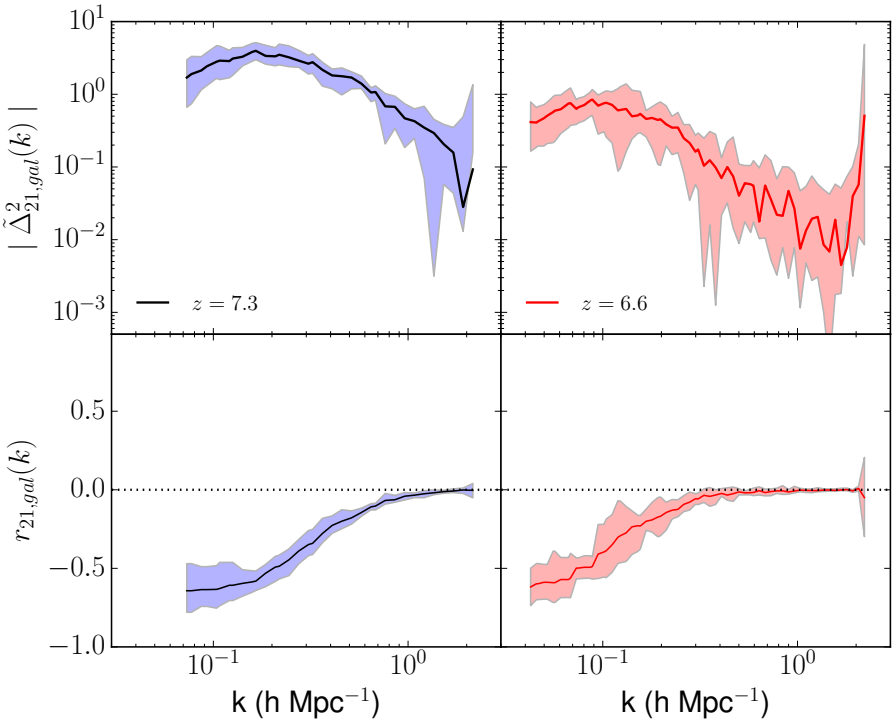

Figure 2.3: Top panels: 2D unnormalized by δT_{b0}, circularly averaged 21cm-galaxy cross-power spectra at $z = 7.3$ (left panel) and 6.6 (right). Shaded areas indicate scatter from 10 mock observations. Bottom panels: 21cm-galaxy cross-correlation coefficient, $r_{21,gal}$, corresponding to $\Delta^2_{21,gal}$. The field of view is 1.7 deg^2 and 7 deg^2 at $z = 7.3$ and $z = 6.6$, respectively.

When selecting LAEs for our mock observations I assign intrinsic Lyα equivalent widths (EW) to the galaxy sample according to a log-normal distribution, as was done by [55]. The distribution is designed to approximately fit observations made by [57], while giving 65 % of the galaxies EW below 25 Å (consistent with [110], as shown in Fig. 1 in [55]). Only galaxies with EW > 0 Å are LAEs. I first selected all LAEs with EW > 20 Å (consistent with HSC expectations), and among these only 1375 (20) most luminous ones at $z = 6.6$ (7.3), to match the number expected to be observed by HSC.

[4]I note that while the solid lines represent the absolute value of the average cross-power spectrum (i.e. the average could be both positive and negative), the shaded area is obtained from the scatter in absolute averaged values (i.e. only positive numbers). For this reason the solid lines do not always lie at the center of the shaded areas.

LOFAR will be detecting the cosmological 21cm signal with a field of view of 5×5 deg^2, and an angular resolution of $3.5'$ [128]. To simulate the LOFAR noise at each frequency, a LOFAR measurement set was filled (the real and imaginary parts of the visibilities) with Gaussian random numbers. This was then imaged (Fourier transformed, accounting for the proper weighting) to obtain noise maps in real-space, and their root mean square was normalised according to (e.g. [112]):

$$\sigma_n = \frac{W}{\eta_s} \frac{\text{SEFD}}{\sqrt{2N(N-1)\Delta\nu t_{int}}}, \tag{2.4}$$

where W is a factor used to increase the noise according to the adopted weighting scheme, η_s is the system efficiency, SEFD is the system equivalent flux density, N is the number of stations, $\Delta\nu$ is the bandwidth, and t_{int} is the integration time. Based on empirical SEFD values for LOFAR (e.g. 3000 Jy at 150 MHz towards the zenith; [116]), I expect σ_n to be about 76 mK at a resolution of 3.5 arcmin, at 150 MHz, after 600 hours and 0.5 MHz of integration and assuming $N = 48$, $W = 1.3$, $\eta_s = 0.9$. Note that adopted noise values are indicative only, and they may change in the actual observations due to e.g. time-variable station projection losses of sensitivity, smaller system efficiency, etc. ([116]). More details about simulating the LOFAR noise can be found in e.g. [89]. The simulated LOFAR noise was added to the brightness temperature map from the simulation.

In Fig. 2.4 I plot the resulting 2D unnormalized by δT_{b0}, circularly averaged 21cm-LAE cross-power spectra with (solid lines) and without (dashed) LOFAR noise. Despite the spectra being much noisier than the previous ones at all scales, a dependence of the normalization on redshift (i.e. amount of HI) and an anti-correlation ($r_{21,gal} \approx -0.30$ (-0.20) at $z = 6.6$ (7.3)) are still visible on large scales, with levels of significance [5] of $p = 0.003$ at $z = 6.6$ and $p = 0.08$ at $z = 7.3$, although the turnover point cannot be clearly identified. Even without LOFAR noise, observations at small scales will still be largely affected by shot noise and will not offer any reliable data (see sec. 2.4). From this analysis I conclude that only scales larger than ~ 60 (45) h^{-1} Mpc, i.e. $k < 0.1$ (0.14) h Mpc^{-1}, at $z = 6.6$ (7.3) can be used for cross-correlation studies.

For a HSC FoV equal to the one of LOFAR, though, I would expect the detection of 3140 and 90 LAEs at $z = 6.6$ and 7.3, respectively. In this case (see Fig. 2.5) the overall noise would be reduced, the anti-correlation signal would be stronger ($r_{21,gal} \approx -0.50$ (-0.30) at $z = 6.6$ (7.3)), large scales could be more reliably used, and information could be extracted down to ~ 60 (30) h^{-1} Mpc, i.e. $k > 0.1$ (0.2) h Mpc^{-1}, at $z = 6.6$ (7.3). In addition, also information at scales larger than ~ 130 (80) h^{-1} Mpc, i.e. $k < 0.05$ (0.08) h Mpc^{-1}, at $z = 6.6$ (7.3) and up to ~ 310 h^{-1} Mpc, i.e. $k \sim 0.02$ h Mpc^{-1}, would be available.

[5] The level of significance, or p-value, is the probability of obtaining at least as extreme result given that the null hypothesis is true. The null hypothesis in this case is that the two fields are not correlated, $r = 0$. It is calculated from the cross-correlation coefficient, r, as $t = (r\sqrt{n-2})(1-r^2)^{-0.5}$, where n is the sample size. t gives us the position of the result in the normal distribution from which the p-value is calculated. Results are usually considered significant if $p < 0.05$.

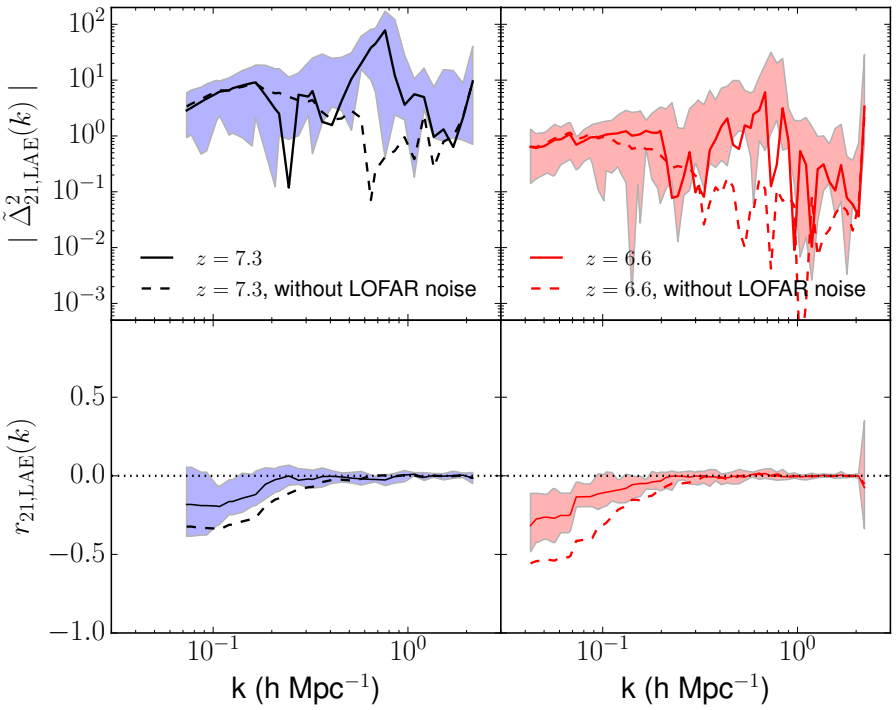

Figure 2.4: Top panels: 2D unnormalized by δT_{b0}, circularly averaged 21cm-LAE cross-power spectra at $z = 7.3$ (left panel) and 6.6 (right). Shaded areas indicate scatter from 10 mock observations. Solid (dashed) lines refer to the cross-power spectra with (without) LOFAR noise. Bottom panels: 21cm-LAE cross-correlation coefficient, $r_{21,\mathrm{LAE}}$, corresponding to $\Delta^2_{21,\mathrm{LAE}}$. The field of view is 1.7 deg^2 and 7 deg^2 at $z = 7.3$ and $z = 6.6$, respectively.

2.2.2 Cross-correlation function

In this section I present my calculations of the theoretical and observational cross-correlation functions.

Theoretical 21cm-galaxy cross-correlation function

Contrary to the cross-power spectrum, the cross-correlation function shows how the correlation between two fields changes in real space. The cross-correlation function between fields a and b is defined as $\xi_{a,b}(\mathbf{r}) = \langle \delta_a(\mathbf{x})\delta_b(\mathbf{x}+\mathbf{r})\rangle$, where $\delta(\mathbf{x})$ is the fractional fluctuation of the field at location $\mathbf{x}$.

The 3D 21cm-galaxy cross-correlation function can then be calculated from the cross-

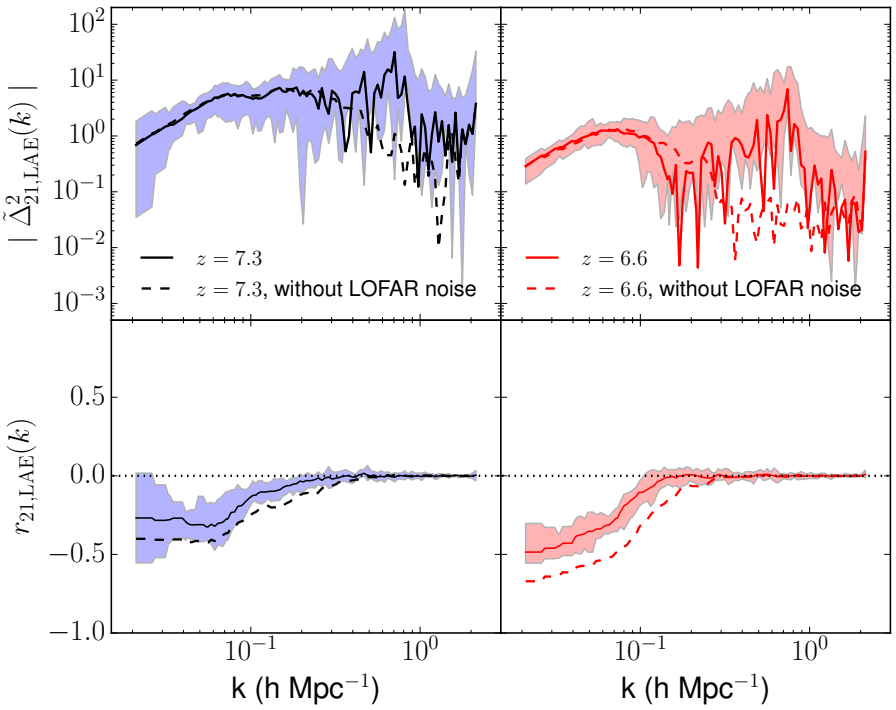

Figure 2.5: Same as Fig. 2.4, but for a HSC field of view of 16 deg^2 at both redshifts.

power spectrum as [87][6]:

$$\xi_{21,gal}(r) = \frac{1}{(2\pi)^3} \int P_{21,gal}(k) \frac{\sin kr}{kr} 4\pi k^2 dk. \tag{2.5}$$

In Fig. 2.6 I show the 21cm-galaxy cross-correlation function at $z = 7.3$, together with the different terms that contribute to it. $\xi_{\rho,gal}$ shows positive correlation on small scales, where there is an overdensity of both gas and galaxies, and no correlation on large scales. The neutral and galaxy fields, $\xi_{x_{HI},gal}$, are anti-correlated on small scales (where the gas is mostly ionized and there is an overdensity of galaxies), mildly correlated on scales just larger than the typical scale of ionized bubbles (where neutral hydrogen is more abundant), and show no correlation on large scales (where most of the neutral hydrogen resides, but there is a paucity of galaxies). $\xi_{x_{HI}\rho,gal}$ and $\xi_{21,gal}$ behave similarly to $\xi_{x_{HI},gal}$, although $\xi_{x_{HI}\rho,gal}$ turns over to positive values and no correlation on much smaller scales. The typical scale of ionized regions is $\sim 50\ h^{-1}$ Mpc.

In Fig. 2.7 I show the theoretical 3D 21cm-galaxy cross-correlation function at $z = 6.68$, 7.06, 7.3, and 8.06. The qualitative behaviour of the curves is similar, with an anti-

[6]Note that when computing the cross-correlation function from the cross-power spectrum, uncertainties arise because of the integration over a finite box size and finite resolution (e.g. uncertainties in the information about the turn over scale; [87]). However, because of the large box size and number of galaxies, it is computationally much more efficient to compute the 3D 21cm-galaxy cross-correlation function from the cross-power spectrum than directly.

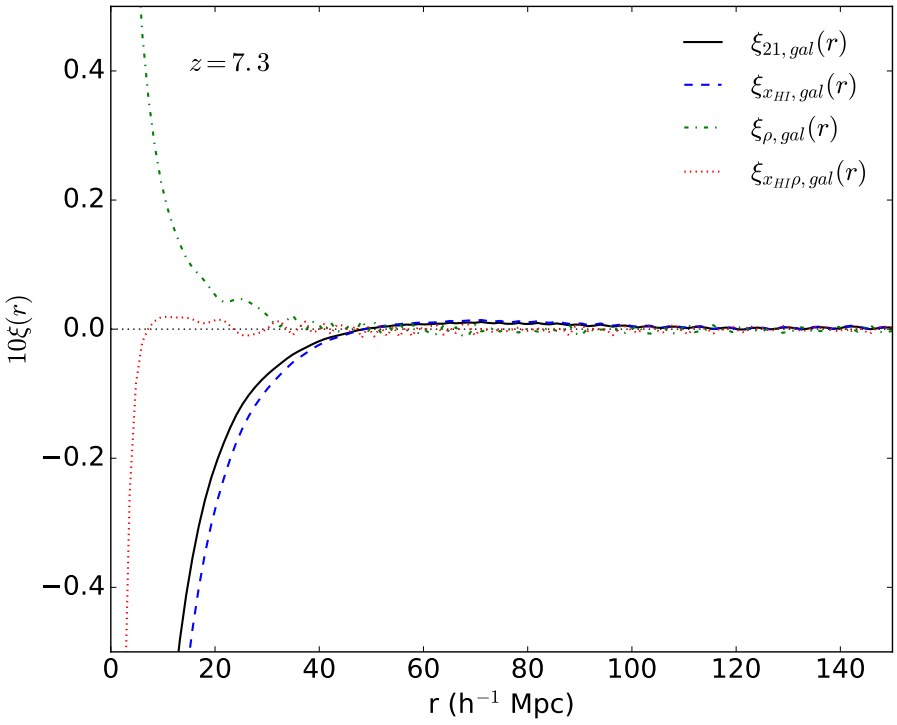

Figure 2.6: Theoretical 3D cross-correlation functions at $z = 7.3$ multiplied by 10 for better resolution: 21cm-galaxy, $\xi_{21,gal}$ (black solid line), neutral fraction-galaxy, $\xi_{x_{HI},gal}$ (blue dashed), density-galaxy, $\xi_{\rho,gal}$ (green dashed-dotted) and neutral density-galaxy, $\xi_{x_{HI}\rho,gal}$ (red dotted). The black dotted line indicates zero correlation.

correlation on small scales, indicating the typical scale of the ionized regions, followed by a small positive correlation, and no correlation on larger scales. As for the case of the power spectrum, the anti-correlation is smaller with decreasing redshift due to the fainter 21cm signal.

Observed 21cm-LAE cross-correlation function

The observed 2D cross-correlation function can be calculated as (e.g. [19]):

$$\xi_{21,LAE}(r) = \frac{\sum_{\mathbf{x}} \delta_{LAE}(\mathbf{x})\delta_{21}(\mathbf{x}+\mathbf{r})}{N_{pair}(r)}, \tag{2.6}$$

where δ_{LAE} and δ_{21} are fractional fluctuations of the LAE and 21cm fields, respectively, and $N_{pair}(r)$ is the number of 21cm-LAE pairs at a separation r.

In Fig. 2.8 I plot the 2D 21cm-LAE cross-correlation functions for our mock observations with a field of view of 1.7 deg^2 at $z = 7.3$ and of 7 deg^2 at $z = 6.6$ (i.e. the equivalent of Fig. 2.4), together with those expected for a larger field of view of 16 deg^2 (i.e. the equivalent of Fig. 2.5). The observed cross-correlation functions show a behavior

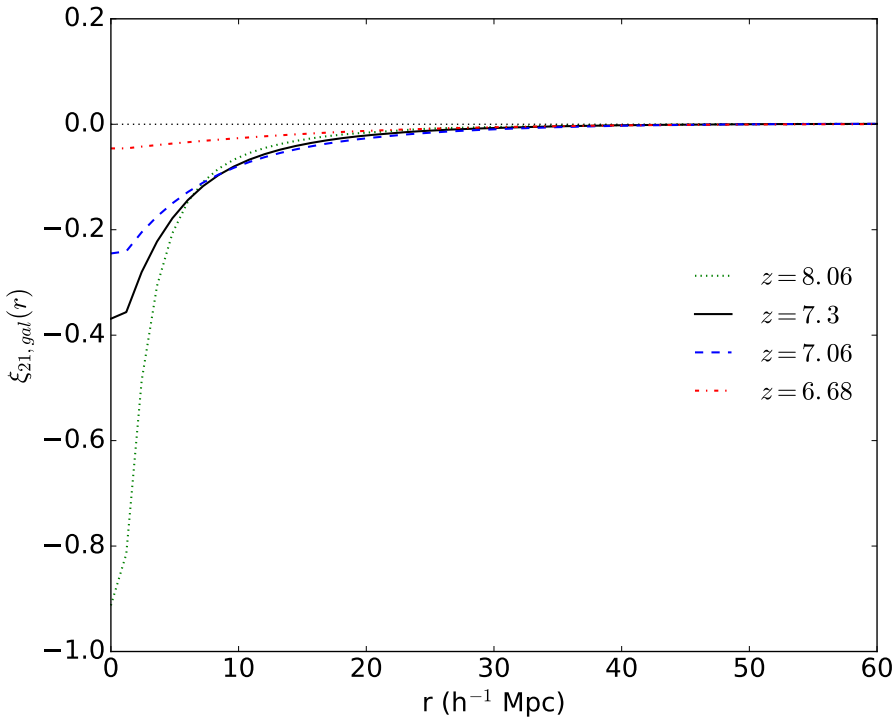

Figure 2.7: Theoretical 3D 21cm-galaxy cross-correlation function at $z = 6.68$ (red dashed-dotted line), 7.06 (blue dashed), 7.3 (black solid), and 8.06 (green dotted). The black dotted line indicates zero correlation.

similar to the theoretical ones. Noise is large at all scales, resulting in a large scatter. While the average of 10 mock observations for both redshifts shows clear anti-correlation at small scales which goes towards no correlation at large scales, scatter is large, so the detection of the anti-correlation might not be possible in a single mock observation. The anti-correlations become much clearer in larger fields of view.

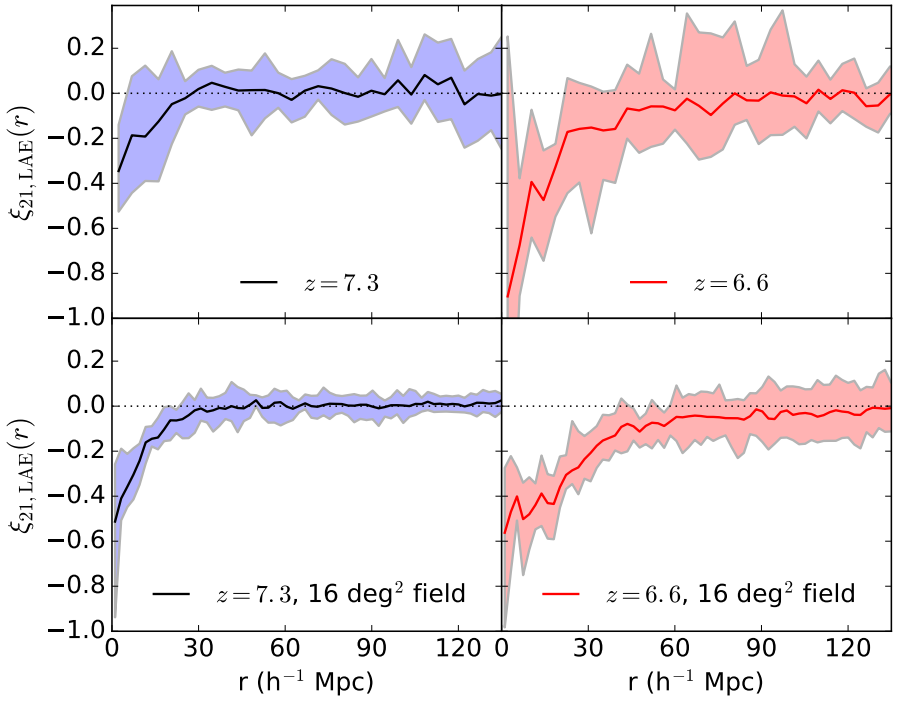

Figure 2.8: 2D 21cm-LAE cross-correlation function for our mock observations with field of view of 1.7 deg^2 at $z = 7.3$ (top left panel) and 7 deg^2 at $z = 6.6$ (top right), and for mock observations with fields of view of 16 deg^2 at $z = 7.3$ (bottom left) and $z = 6.6$ (bottom right). The black dotted lines indicate zero correlation and shaded areas indicate scatter from 10 mock observations.

2.3 21cm-galaxy cross-correlations observable with SKA and future galaxy surveys

In this section I will show the 21cm-galaxy cross-correlations as they would be observable with the next generation radio array SKA and future galaxy surveys. The galaxy surveys that are going to be considered here are Subaru HSC, Subaru PFS, and WFIRST.

The 21cm mock observations of SKA observations are built by adding the SKA noise to the brightness temperature maps. The instrumental effects were modelled using the OSKAR simulator[7] and a list of preliminary station positions for the SKA1-LOW. Here I follow the procedure as explained in [13]. The images of the SKA1-LOW point spread function (PSF) were produced by assuming a full correlation between all 866 core stations. The maximum baseline length is 5.29 km. The telescope was located at 52.7 degrees latitude and baseline coordinates were generated for a 12-hour synthesis observation and a phase centre on the sky at apparent equatorial coordinates $(\alpha, \delta) = (218, 34.5)$ degrees. Each of 374545 baselines got 144 snapshots in a 5-minute sampling interval. Common Astronomy

[7] http://www.oerc.ox.ac.uk/∼ska/oskar

Software Applications package [9] was then used to generate PSF images across a 5-degree field-of-view using 256 w-projection planes. A 1000 hour integration time was used for the noise normalization using the prescription described in e.g. [113].

2.3.1 Subaru Hyper Supreme Cam

Here I follow the same approach as in the previous section. I reduce the simulation box to match the size of HSC's FoV (7 deg^2 at $z = 6.6$ and 1.7 deg^2 at $z = 7.3$), depth ($\Delta z = 0.1$), and expected number of observed LAEs per field (1375 at $z = 6.6$ and 20 at $z = 7.3$) and then collapse the reduced sub-box to a single plane to mimic HSC's observations. In this process 10 mock observations are created. To build SKA mock observation I also reduce the simulation box in the same way (to have FoV of 7 deg^2 at $z = 6.6$ and 1.7 deg^2 at $z = 7.3$, and depth of $\Delta z = 0.1$), but I add SKA noise instead of LOFAR noise to the brightness temperature maps.

I cross-correlate a HSC mock observation with each slice in the corresponding SKA mock observation and then average the resulting 21cm-LAE cross-power spectra (or cross-correlation functions) to get a result from a single FoV. From the 10 mock observations I obtain 10 21cm-galaxy cross-correlations which are then again averaged to get a sample independent result.

The resulting 2D unnormalized by δT_{b0} 21cm-LAE cross-power spectra are shown in Fig. 2.9. Comparing the resulting cross-power spectra with the ones between LOFAR and HSC (Fig. 2.4) it is easy to notice that the spectra observed by SKA and HSC are smoother, more similar to the theoretical ones, and with smaller scatter. The 21cm-LAE cross-power spectra with and without the SKA noise show very little difference, and the resulting cross-correlation coefficients show stronger anti-correlation (~ -0.4 at $z = 7.3$ and ~ -0.6 at $z = 6.6$) than the cross-correlation coefficients observed with LOFAR and HSC.

The 21cm-LAE cross-correlation functions are computed following Eq. 2.6 and are shown in Fig. 2.10. The 21cm-LAE cross-correlation functions also look much less noisy when observed with SKA instead of LOFAR (Fig. 2.8), with smaller scatter and clearer anti-correlation on small scales. The difference between graphs with or without SKA noise is minimal.

2.3.2 Subaru Prime Focus Spectrograph

After Subaru finishes observing with HSC, the HSC will be replaced with the PFS which will observe the same FoV as HSC and will focus on the previously detected LAEs, but will give precise positions for the observed LAEs due to spectroscopic observations. Hence, a 3D instead of 2D cross-correlations can be measured from its observations.

I reduce the simulation box to match the size of a PFS's FoV (7 deg^2 at $z = 6.6$ and 1.7 deg^2 at $z = 7.3$), depth ($\Delta z = 0.1$), and expected number of observed LAEs per field (1375 at $z = 6.6$ and 20 at $z = 7.3$), but unlike for HSC's mock observations I do not collapse the reduced sub-box to a single plane, since PFS will acquire the exact positions for the

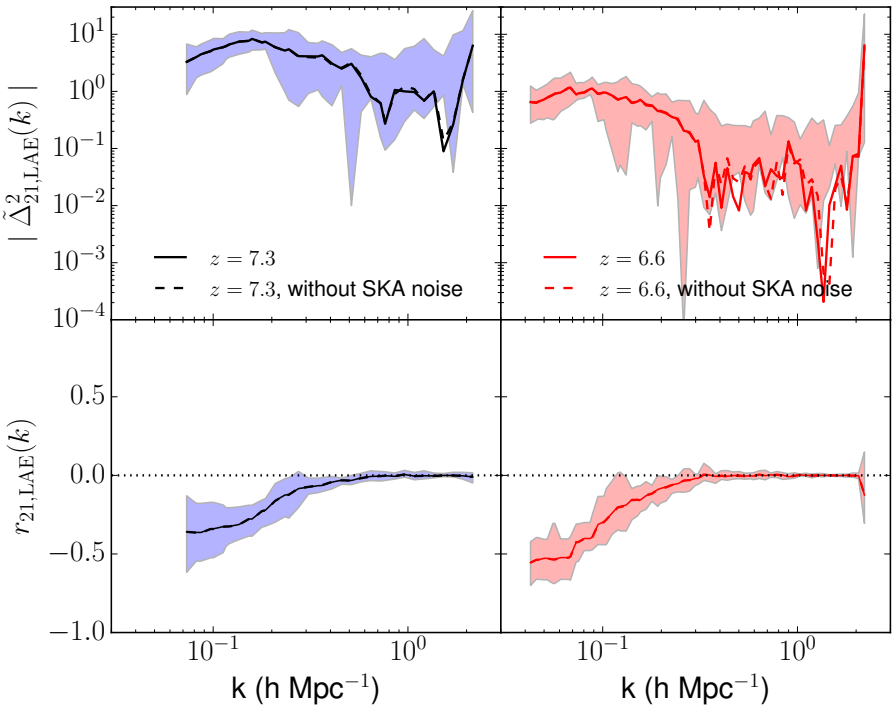

Figure 2.9: Top panels: 2D unnormalized by δT_{b0}, circularly averaged 21cm-LAE cross-power spectra between SKA and HSC at $z = 7.3$ (left panel) and 6.6 (right). Shaded areas indicate scatter from 10 mock observations. Bottom panels: 21cm-LAE cross-correlation coefficient, $r_{21,\mathrm{LAE}}$, corresponding to $\Delta^2_{21,\mathrm{LAE}}$. The field of view is 1.7 deg^2 and 7 deg^2 at $z = 7.3$ and $z = 6.6$, respectively.

observed LAEs. From the given reductions of the simulation box 10 3D mock observations are created. SKA mock observations are created as explained in the previous sub-section. I then cross-correlate each LAEs mock-observation with a 21cm mock observation, and in this process I get 10 cross-correlations which are then averaged to get a sample independent result.

The resulting 3D, unnormalized, spherically averaged 21cm-LAE cross-power spectra are shown in Fig. 2.11. They look very similar to the ones measured with SKA and HSC. However, the strength of the cross-power spectra is now lower due to the spherical average over a 3D volume, where the reduction of strength is higher at $z = 6.6$ due to the larger observed volume. The cross-power spectra become noisy around the same scale as the cross-power spectra measured with SKA and HSC, since they have similar shot noise (due to the same number of LAEs). The strength of the anti-correlation at large scales in the cross-correlation coefficient is similar to the ones measured with SKA and HSC, ~ -0.4 at $z = 7.3$ and ~ -0.6 at $z = 6.6$. However, there is less scatter in the cross-correlation coefficients at $z = 7.3$.

The cross-correlation function between SKA and PFS shows much more scatter than the one between SKA and HSC and a weaker anti-correlation between the two fields, with

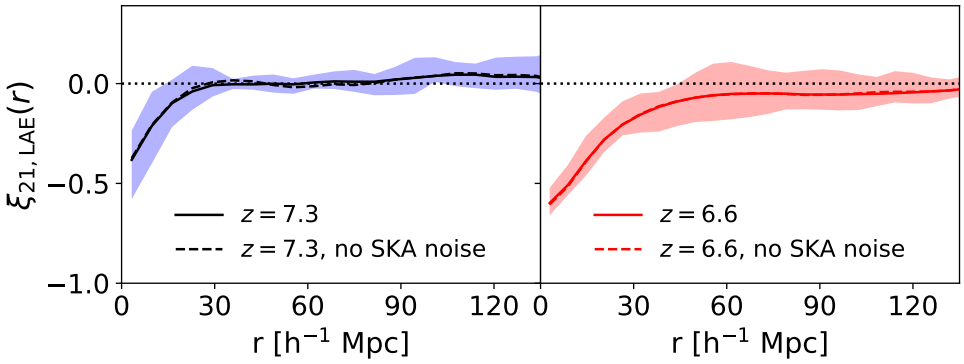

Figure 2.10: 2D 21cm-LAE cross-correlation function between SKA and HSC with field of view of 1.7 deg^2 at $z = 7.3$ (left panel) and 7 deg^2 at $z = 6.6$ (right). The black dotted lines indicate zero correlation and shaded areas indicate scatter from 10 mock observations.

the anti-correlation reaching -0.2 at $z = 7.3$ and -0.1 at $z = 6.6$ (Fig. 2.12). The increase in scatter and decrease in the anti-correlation are due to the attempt of describing a 3D space with only 20 (1375) LAEs at $z = 7.3$ (6.6). The results without the SKA noise do not improve much since this effect is due to the large shot noise.

2.3.3 Wide Field Infrared Survey Telescope

WFIRST is a space telescope that will be observing high redshift LAEs at redshift range $7.5 \leq z \leq 8.5$ with the expectation to observe 900 LAEs/deg^2/z with EW$>$ 5 Å. Observations will be done spectroscopically over wide areas of the sky. I constructed a mock observation for WFIRST observation with a field of view of our simulation box (16 deg^2), a full redshift in depth and 14400 observed LAEs, as expected to be observed with WFIRST in this volume. The SKA mock observation was created by adding the SKA noise to the brightness temperature maps and it was designed to have the same field of view and depth. These mock observations were then cross-correlated to give a 3D 21cm-LAE cross-power spectrum and cross-correlation function.

The resulting 21cm-LAE cross-power spectrum is shown in Fig. 2.13. I show that the 21cm-LAE cross power spectrum has a very similar shape to the theoretically expected shape (Fig. 2.2) but it becomes noisy on small scales, where $k > 1$ h Mpc^{-1}. The corresponding cross-correlation coefficient shows strong anti-correlation on large scales ($r_{21,LAE} \sim -0.9$). As previously, SKA noise has very little influence on the resulting cross-power spectrum and cross-correlation coefficient.

Fig. 2.14 shows 21cm-LAE cross-correlation function with and without SKA noise. The difference between the results is very small, and the main difference is a small decrease in the amplitude at small scales, due to the addition of the SKA noise, from -1 to ~ -0.9. The addition of the SKA noise does not otherwise affect the shape of the cross-correlation function. The cross-correlation function shows that the average scale of ionized bubbles at the redshift in question is $\sim 30h^{-1}$ Mpc.

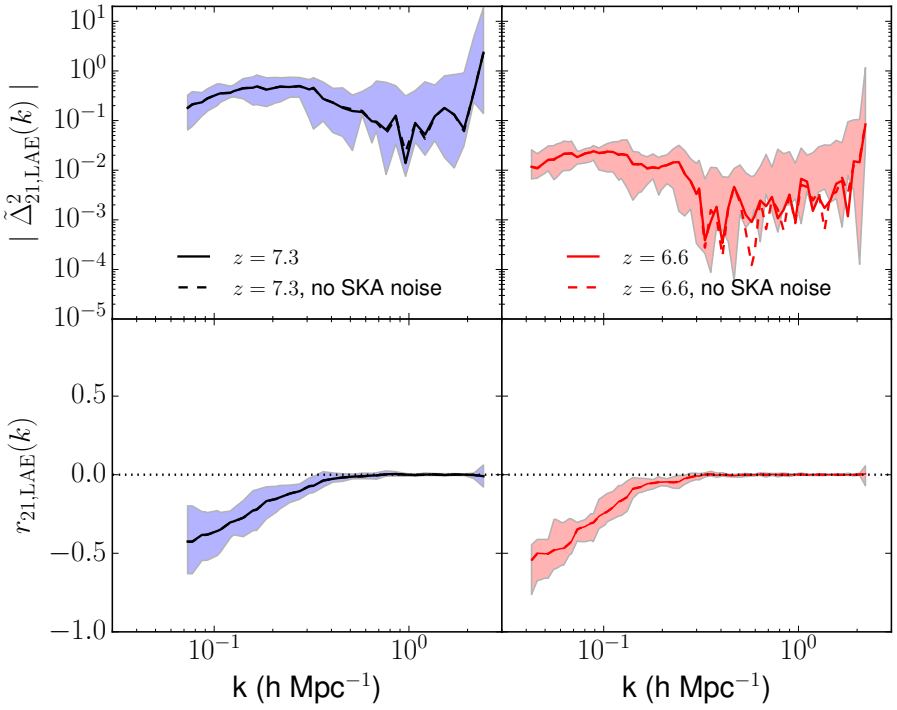

Figure 2.11: Top panels: 3D unnormalized by δT_{b0}, spherically averaged 21cm-LAE cross-power spectra between SKA and PFS at $z = 7.3$ (left panel) and 6.6 (right). Shaded areas indicate scatter from 10 mock observations. Bottom panels: 21cm-LAE cross-correlation coefficient, $r_{21,\mathrm{LAE}}$, corresponding to $\Delta^2_{21,\mathrm{LAE}}$. The field of view is 1.7 deg^2 and 7 deg^2 at $z = 7.3$ and $z = 6.6$, respectively.

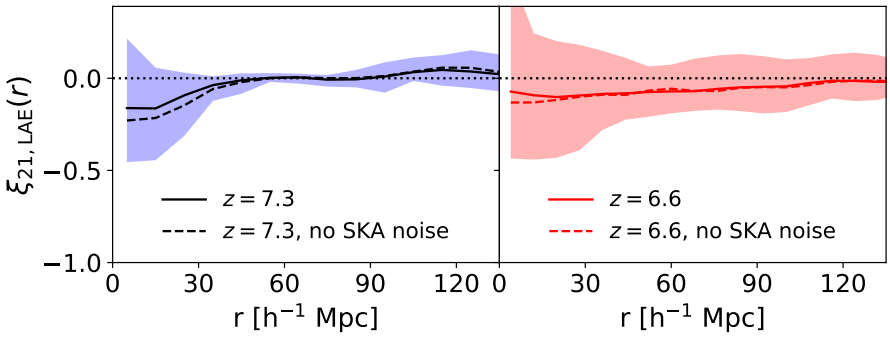

Figure 2.12: 3D 21cm-LAE cross-correlation function between SKA and PFS with field of view of 1.7 deg^2 at $z = 7.3$ (left panel) and 7 deg^2 at $z = 6.6$ (right). The black dotted lines indicate zero correlation and shaded areas indicate scatter from 10 mock observations.

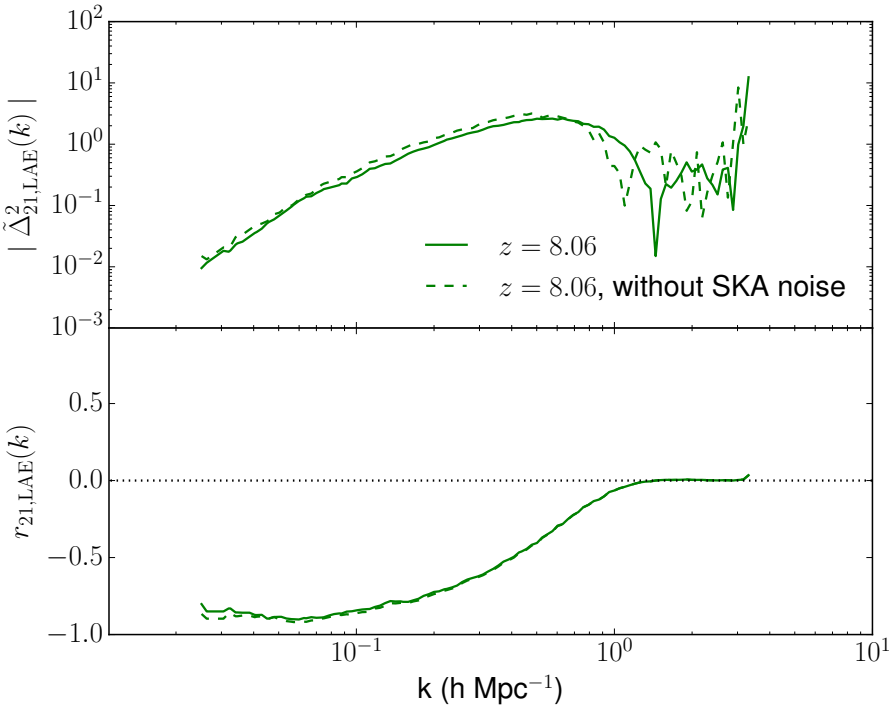

Figure 2.13: Top panel: 3D, unnormalized by δT_{b0}, spherically averaged 21cm-LAE cross-power spectrum between SKA and WFIRST at $z = 8.06$. Solid (dashed) line refers to the cross-power spectrum with (without) SKA noise. Bottom panel: 21cm-LAE cross-correlation coefficient, $r_{21,\mathrm{LAE}}$, corresponding to $\Delta^2_{21,\mathrm{LAE}}$. The field of view is 16 deg^2.

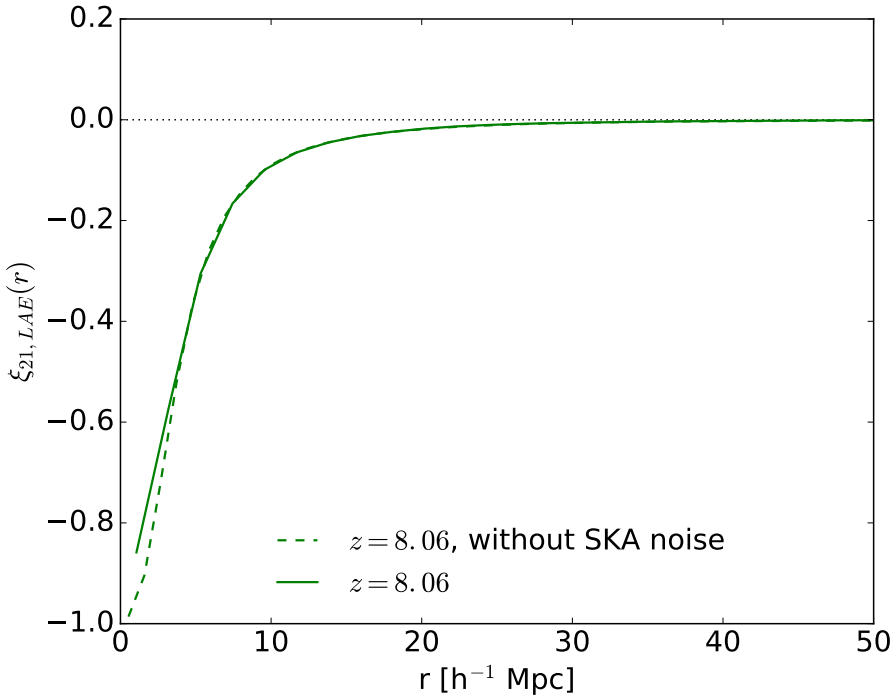

Figure 2.14: 3D 21cm-LAE cross-correlation function between SKA and WFIRST with field of view of 16 deg^2 at $z = 8.06$ with (full green line) and without (dashed green) SKA noise. The black dotted line indicates zero correlation.

2.4 Discussion

Observations of 21cm emission and high-z LAEs are extremely challenging, and both will suffer from severe noise problems. Even assuming that foregrounds subtraction will work perfectly, the system noise will still largely exceed the expected signal, in particular at the smaller scales, so that possibly only scales larger than $\sim 60\ h^{-1}$ Mpc (corresponding to $k \sim 0.1h$ Mpc^{-1}) will be accessible by a telescope like LOFAR. In addition, the field of view of HSC is much smaller than that of LOFAR, so that only a fraction of the large scales observed by LOFAR will be covered also by HSC.

To illustrate this issue further, in Fig. 2.15 I show the 2D 21cm auto-power spectra with and without the LOFAR noise after 600 hours of observation in fields of view of 7 deg^2 at $z = 6.6$ and 1.7 deg^2 at $z = 7.3$, i.e. equivalent to the ones of HSC[8]. At both redshifts noise on scales smaller than $\sim 60h^{-1}$ Mpc (i.e. $k \sim 0.1h$ Mpc^{-1}) is orders of magnitude larger than the expected signal, while it decreases gradually at larger scales. Noise on large scales at $z = 6.6$ is somewhat larger than at $z = 7.3$.

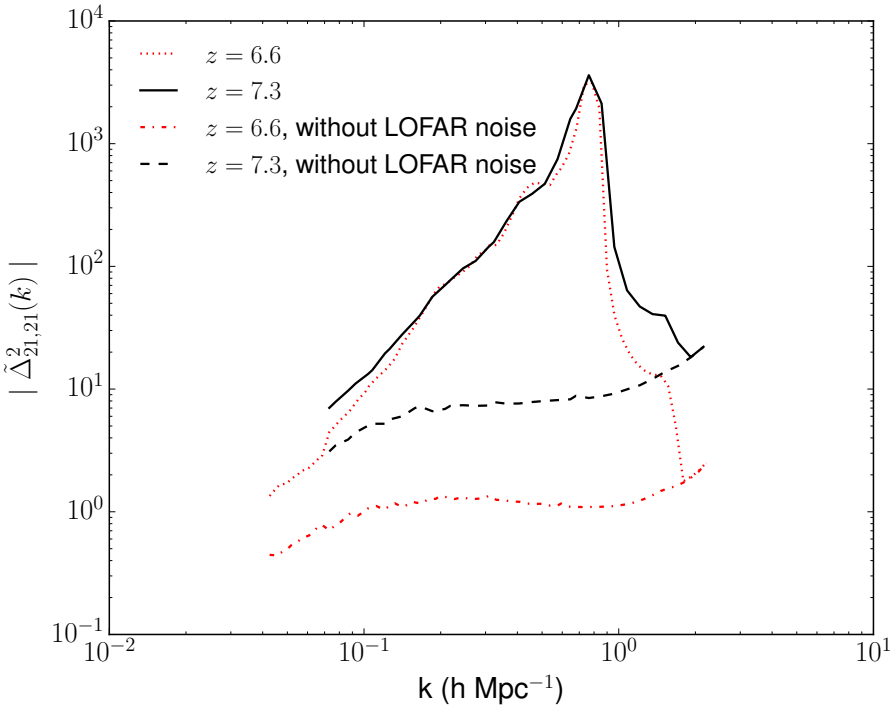

Figure 2.15: 2D, unnormalised by δT_{b0}, 21cm auto-power spectrum with (upper set of curves) and without (lower set) LOFAR noise. The field of view is 1.7 deg^2 and 7 deg^2 at $z = 7.3$ and $z = 6.6$, respectively.

[8]Note that in observations of the 21cm auto-power spectra the expectation value of the noise power spectrum can be subtracted from the measurements. However, in observations of the 21cm-LAE cross-power spectra this is not possible since instrumental effects from LAE observations are also present and their influence cannot be treated separately. The same reasoning applies to LAE observations.

The HSC observations discussed here are groundbreaking, as they will increase the number of detected high-z LAEs by at least one order of magnitude. However, substantial shot noise is still expected, as shown in Fig. 2.16. Observations at both redshifts will be dominated by shot noise at scales below $10h^{-1}$ Mpc (i.e. $k > 0.6h$ Mpc^{-1}) at z=6.6, and $30h^{-1}$ Mpc (i.e. $k > 0.2h$ Mpc^{-1}) at z=7.3, while at large scales the LAEs auto-power spectrum is stronger than that of the shot noise, in particular at $z = 6.6$.

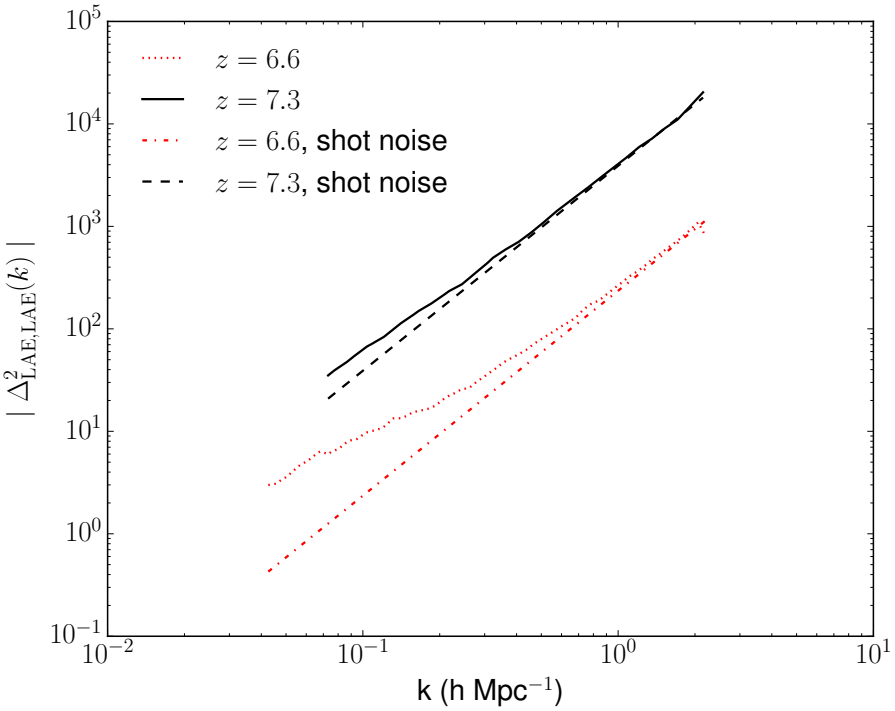

Figure 2.16: 2D LAE auto-power spectra at $z = 7.3$ (black solid line) and $z = 6.6$ (red dotted), and shot noise power spectra at $z = 7.3$ (black dashed) and $z = 6.6$ (red dashed-dotted). The field of view is 1.7 deg^2 and 7 deg^2 at $z = 7.3$ and $z = 6.6$, respectively.

Since the 21cm-LAE cross-power spectrum observed with LOFAR and HSC will be affected by noise from both instruments, I expect to be able to probe only scales larger than $\sim 60h^{-1}$ Mpc (i.e. $k < 0.1h$ Mpc^{-1}). Such scales will still have shot noise, in particular at $z = 7.3$, but this should not prevent the detection of an anti-correlation.

Stronger anti-correlation could be detected by reducing the noise, e.g. increasing the integration time for 21cm observations ($\sigma_{noise} \sim t_{int}^{-1/2}$ for LOFAR; Eq. 2.4), or with a larger field of view. The latter would increase the number of observed LAEs and thus reduce the shot noise and extend the number of observed k-modes.

In 21cm-LAE cross-correlation function observed with LOFAR and HSC, noise is large at all scales. This is because, unlike in the cross-power spectrum, noise does not get separated by its k-modes, and thus it is equally distributed on all scales. Comparing cross-correlation functions at $z = 7.3$ for our reference mock observations and for mock

observations with a larger field of view (Fig. 2.8), in the latter case a larger amplitude of the anti-correlation as well as a smaller scatter can be observed because of a reduction of the noise component. While the scatter is smaller also at $z = 6.6$, the amplitude is not increased. I suggest that this is due to the LOFAR noise, which has a stronger effect at $z = 6.6$. The noise should also be responsible for the positive correlation observed at large scales in the top panels of Figures 2.17 and 2.18.

To reduce the noise levels I also smoothed the 21cm field with a Gaussian filter with standard deviations of $\sigma = 2$, 5 and 10 (with smoothing radii 4.14 h^{-1} Mpc, 10.35 h^{-1} Mpc and 20.70 h^{-1} Mpc, respectively) at $z = 6.6$ (Fig. 2.17) and $\sigma = 1$, 2 and 5 (with smoothing radii 2.07h^{-1} Mpc, 4.14 h^{-1} Mpc and 10.35 h^{-1} Mpc, respectively) at $z = 7.3$ (Fig. 2.18). It is clear that both the average cross-correlation function and the scatter become smoother with increasing σ. However, by smoothing the field I also loose information (e.g. in terms of anti-correlation amplitude), which is visible when comparing results with different σ. Smoothing the signal at $z = 6.6$ reduces the noise on small scales enough that the anti-correlation becomes clear even for the largest scatter. At $z = 7.3$, instead, the shot noise is larger because of the smaller LAE sample, so that even after smoothing the scatter on small scales remains large.

While smoothing reduces noise in the cross-correlation function (which remains though still noisy on all scales), it is not helpful when applied to the 21cm-galaxy cross-power spectrum. The reason for this is that because the noise is being separated by its k-modes, smoothing would affect only the small scales which would still be over-contaminated by shot noise. However, the separation of noise by its k-modes is exactly what makes large scales observable and the 21cm-galaxy cross-power spectrum a more useful probe of reionization.

The next generation instrument SKA will have much lower noise levels than LOFAR and thus will introduce a significant change when used for the observation of 21cm-LAE cross-correlations at redshifts in question. The 21cm auto-power spectra with and without SKA noise after 1000 hours of observation in fields of view of 7 deg^2 at $z = 6.6$ and 1.7 deg^2 at $z = 7.3$, i.e. equivalent to the ones of HSC and PFS are shown in Fig. 2.19, and in Fig. 2.20 I show 21cm auto-power spectrum with and without SKA noise in the field of view of 16 deg^2 at $z = 8.06$ which corresponds to my mock observations for WFIRST. The difference between results with and without SKA noise is so small it is barely visible. Thus, the main noise contribution to these observations comes from the shot noise due to LAEs observations. The reduction from LOFAR to SKA noise will significantly improve the range of observable scales, from $k < 0.1$ h Mpc^{-1} to $k < 0.3$ h Mpc^{-1}, as well as the detectable strength of the anti-correlation on the large scales of the cross-correlation coefficient, up to ~ -0.4 at $z = 7.3$ and ~ -0.6 at $z = 6.6$.

PFS will introduce a significant change to Subaru's measurements by making precise spectroscopic observations of previously detected LAEs which will enable more precise computations of their positions, thus allowing 3D cross-correlations. However, this important change in LAE information will not result in significant improvement of the measured cross-correlations. This is due to the fact that the same number of LAEs will describe the observed 3D volume poorer than the 2D surface, increasing the shot noise significantly. Thus, gained information about positions of LAEs and the increase of the shot noise will

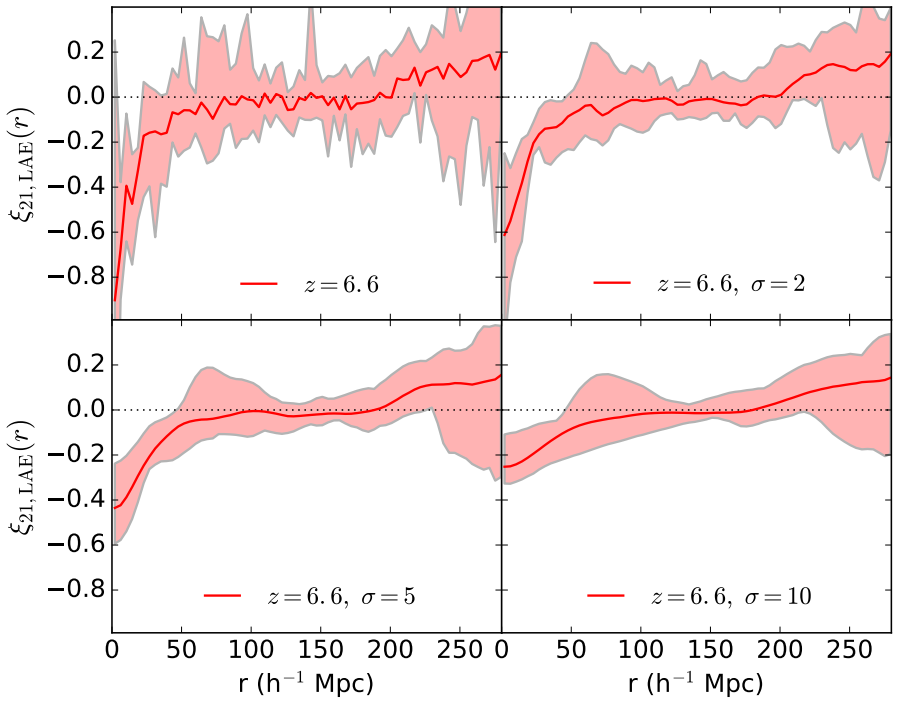

Figure 2.17: 2D 21cm-LAE cross-correlation function at $z = 6.6$ (top left panel) using a Gaussian filter with standard deviation $\sigma = 2$ (top right), $\sigma = 5$ (bottom left) and $\sigma = 10$ (bottom right). The black dotted lines indicate zero correlation and shaded areas scatter from 10 mock observations. The field of view is 7 deg^2.

mostly cancel each other out. As shown in Fig. 2.21, in PFS observations the shot noise is more dominant over the LAE auto-power spectrum than in the HSC's observations (Fig. 2.16). At $z = 7.3$ the LAE auto-power spectrum starts to dominate over shot noise at the same scale as in HSC's observations, however, in PFS's observations this dominance is barely visible and insignificant. At $z = 6.6$ the dominance of the LAE auto-power spectrum starts on somewhat larger scales $k \sim 0.3\ h$ Mpc^{-1}, but the difference between LAE auto-power spectrum and the shot noise reaches only half an order of magnitude. Thus, even though there are no projection effects in PFS's observations, the shape of the cross-power spectra and the strength of the cross-correlation coefficient do not change much due to the simultaneous increase of the shot noise.

In Fig. 2.22 I show LAE auto-power spectrum and shot noise power spectrum for WFIRST observations at $z = 8.06$. The shot noise power spectrum is dominating LAE auto-power spectrum on scales smaller than $k > 1\ h$ Mpc^{-1}, while towards larger scales LAE auto-power spectrum starts dominating the shot noise and on the largest scales it is an order of magnitude larger than the shot noise. Combination of the low shot noise in WFIRST observation and the low SKA noise is giving very precise measurements on scales larger than $k < 1\ h$ Mpc^{-1}, while on smaller scales the shot noise will be too large.

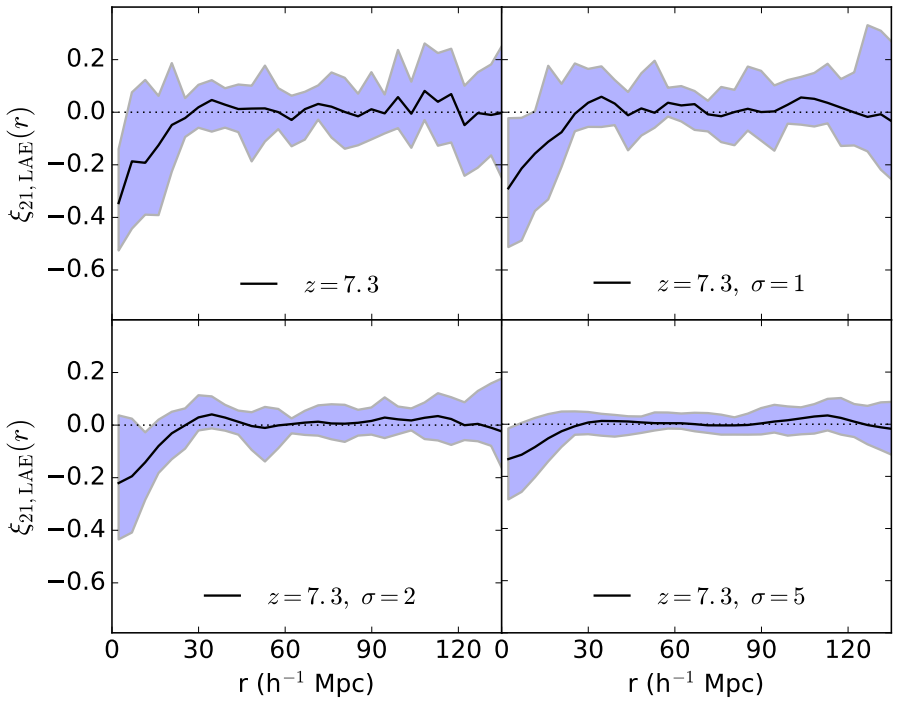

Figure 2.18: 2D 21cm-LAE cross-correlation function at $z = 7.3$ (top left panel) using a Gaussian filter with standard deviation $\sigma = 1$ (top right), $\sigma = 2$ (bottom left) and $\sigma = 5$ (bottom right). The black dotted lines indicate zero correlation and shaded areas scatter from 10 mock observations. The field of view is 1.7 deg^2.

Unlike with LOFAR observations, 21cm-LAE cross-correlation functions observed with SKA and HSC will be much less noisy, smoother, and will have smaller scatter on all scales. However, cross-correlation function between SKA and PFS will be much noisier and with less prominent anti-correlation. This is because of the increase in shot noise due to the increase in space being described with these LAEs when transitioning from 2D to 3D observations. WFIRST will have significantly reduced shot noise, due to observing a large number density of LAEs, resulting in very precise cross-correlation function.

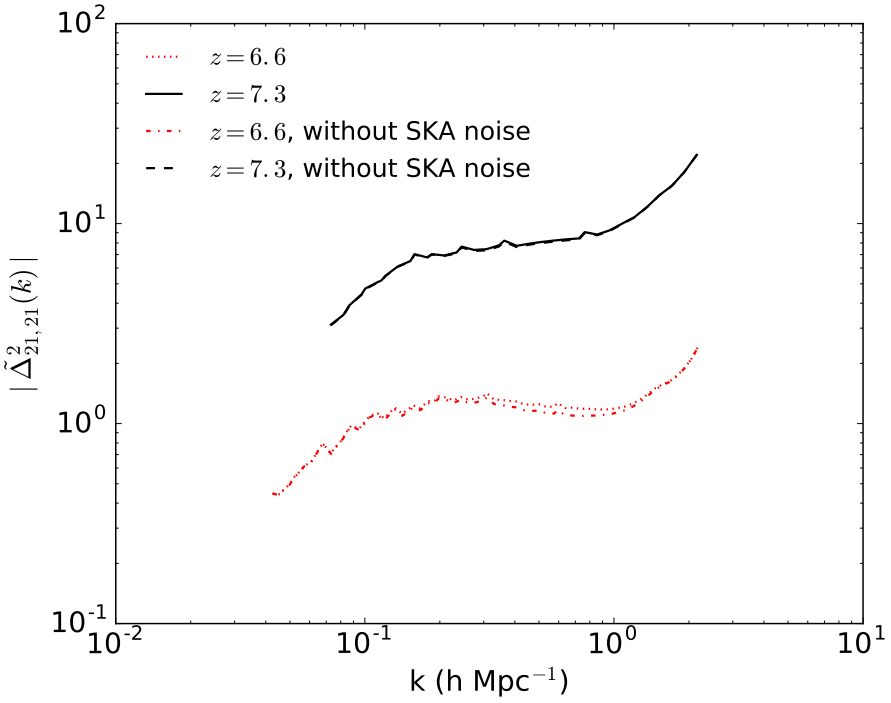

Figure 2.19: 2D, unnormalised by δT_{b0}, 21cm auto-power spectrum with SKA noise at $z = 7.3$ (black solid line) and $z = 6.6$ (red dotted), and without SKA noise at $z = 7.3$ (black dashed) and $z = 6.6$ (red dashed-dotted). The field of view is 1.7 deg^2 and 7 deg^2 at $z = 7.3$ and $z = 6.6$, respectively.

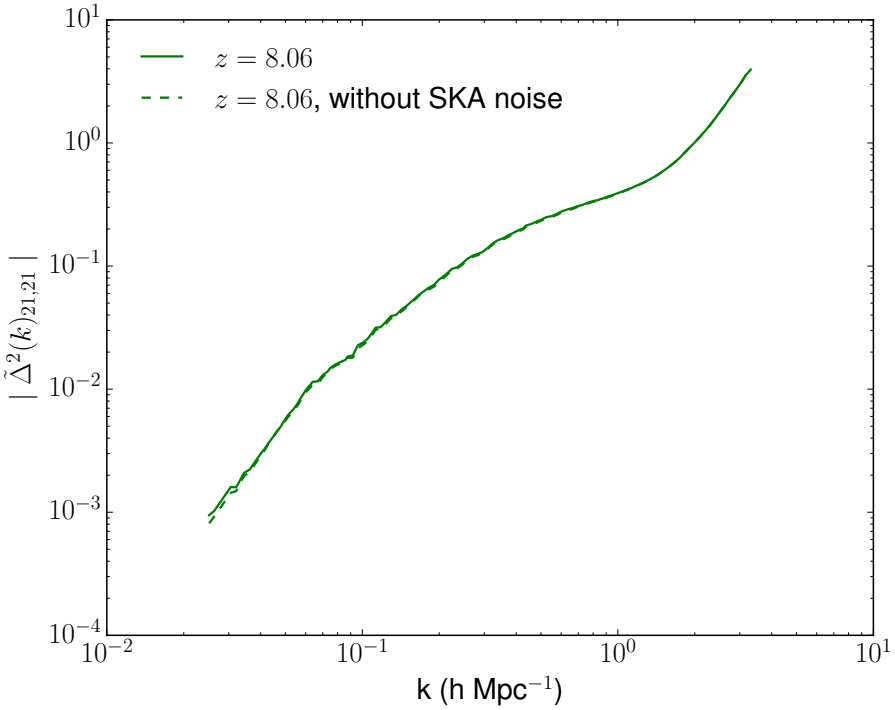

Figure 2.20: 3D, unnormalised by δT_{b0}, 21cm auto-power spectrum with SKA noise (green solid line) and without SKA noise (green dashed) at $z = 8.06$. The field of view is 16 deg^2.

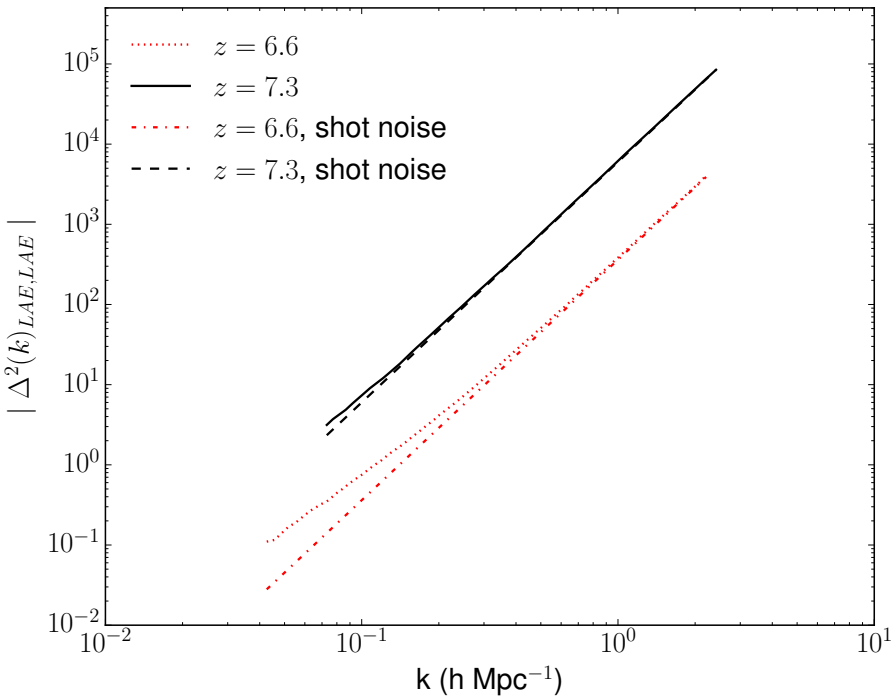

Figure 2.21: 3D LAE auto-power spectra at $z = 7.3$ (black solid line) and $z = 6.6$ (red dotted), and shot noise power spectra at $z = 7.3$ (black dashed) and $z = 6.6$ (red dashed-dotted). The field of view is 1.7 deg^2 and 7 deg^2 at $z = 7.3$ and $z = 6.6$, respectively.

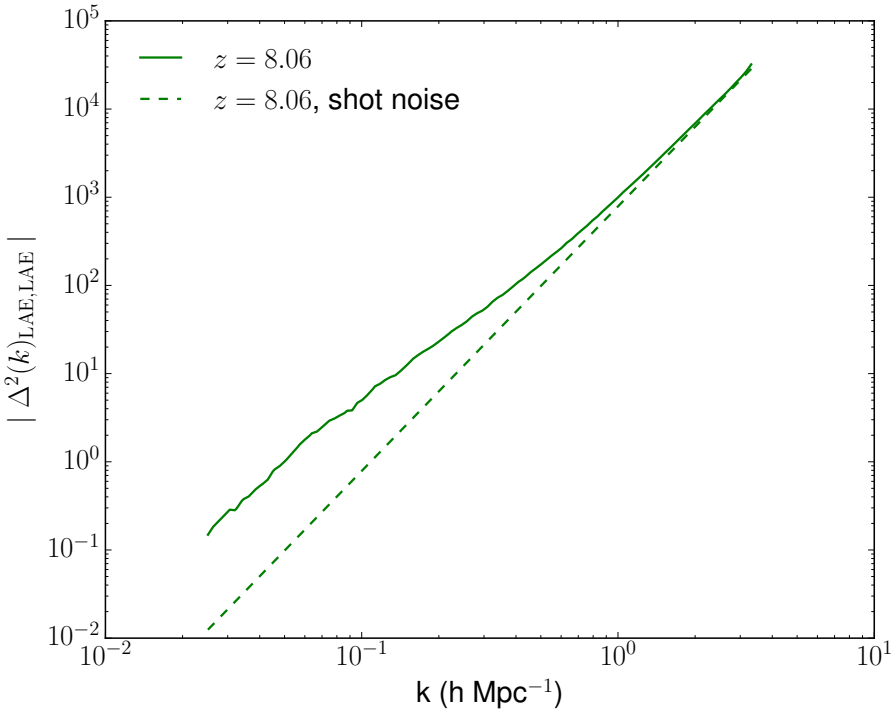

Figure 2.22: 3D LAE auto-power spectrum (green solid line) and shot noise power spectrum (green dashed) at $z = 8.06$. The field of view is 16 deg^2.

2.5 Conclusions

I presented theoretical 3D 21cm-galaxy cross-power spectra at redshift 6.68, 7.06, 7.3, and 8.06 computed from full radiative transfer + N body simulations by [48], which have been post-processed with a Lyα radiative transfer code ([65, 56]) to accurately model the observed properties of LAEs. The mock 21cm and galaxy observed maps obtained by adding the LOFAR and HSC characteristics have been used to compute the 2D 21cm-LAE cross-power spectra and cross-correlation functions for combined LOFAR and HSC observations. The mock 21cm and galaxy observed maps obtained by adding SKA and future galaxy survey characteristics (such as HSC, PFS, and WFIRST) have been used to compute the 2D or 3D 21cm-LAE cross-power spectra and cross-correlation functions for combined observations between SKA and future galaxy surveys. The present investigation improves on previous efforts ([71, 121]) in terms of theoretical modeling (by including e.g. an accurate treatment of the radiative transfer of both line and continuum photons to model the reionization process and the properties of the LAEs) as well as mock observations (by targeting the upcoming observation of the ELAIS-N1 field by the LOFAR and Subaru telescopes, and the future observations of SKA and future galaxy surveys).

Our theoretical 3D 21cm-galaxy cross-power spectra agree with previous investigations (i.e. [71, 121]). More specifically, I am able to recover the same redshift dependence and shape, with a distinct turnover point indicating the typical scale of ionized bubbles. I confirm that the 21cm-galaxy cross-power spectrum could provide information on the progress of reionization and the typical size of HII regions at different redshifts.

The measured 21cm-LAE cross-power spectrum between LOFAR and HSC suffers from projection effects (as it is 2D), as well as from noise in both radio and LAEs detections. LOFAR recently started observations of 21cm emission from neutral hydrogen in the redshift range $z = 6 - 11.4$ [126, 52], while HSC will also soon start its observational campaign with two narrow-band filters searching for LAEs at $z = 6.6$ and 7.3 (M. Ouchi 2012, private conversation). Both telescopes plan to observe the ELAIS-N1 field at $z = 6.6$, making it possible to detect the 21cm-galaxy cross-power spectrum. I constructed mock observations specifically tailored to match LOFAR and HSC campaigns at redshifts 6.6 and 7.3. Our mock observations show that despite the observed spectra being much noisier than the corresponding theoretical 3D ones, dependence of the normalization on redshift (i.e. amount of HI) is clearly visible, as well as the anti-correlation between the two fields, with a cross-correlation coefficient $r_{21,gal} \approx -0.30$ at levels of significance of $p = 0.003$ at $z = 6.6$ and $r_{21,gal} \approx -0.20$ at $p = 0.08$ at $z = 7.3$. However, the turnover point cannot be clearly determined because small scales will be overwhelmed by noise.

I also investigated properties and observability of the 21cm-galaxy cross-correlation functions, which are expected to be negative on small scales, mildly correlated on scales just larger than the typical size of ionized regions, and show no correlation on even larger scales. This agrees well with predictions of the 21cm-galaxy cross-correlation function by [87]. Despite observational effects in LOFAR and HSC measurements, like noise and galaxy number densities, observed correlation functions should retain the theoretical shape. However, unlike the observed 21cm-LAE cross-power spectrum, the correlation function

observed with LOFAR and HSC suffers from a strong noise on all scales (as it does not get separated by its k-modes), thus uncertainties in the signal will be large.

I also explored possible future measurements of cross-correlations between the 21cm fields observed with SKA and the LAEs observed with future galaxy surveys such as Subaru HSC, Subaru PFS, and WFIRST. I show both, 21cm-LAE cross-power spectra and cross-correlation functions, since the noise from these observations affects them differently.

The next generation observations with SKA with much lower noise levels than LOFAR will introduce a significant change when used for the observation of 21cm-LAE cross-correlations with Subaru HSC. The reduction from LOFAR to SKA noise will significantly improve the range of observable scales, from $k < 0.1\ h\ \mathrm{Mpc}^{-1}$ to $k < 0.3\ h\ \mathrm{Mpc}^{-1}$, as well as the detectable strength of the anti-correlation on the large scales of the cross-correlation coefficient, up to ~ -0.4 at $z = 7.3$ and ~ -0.6 at $z = 6.6$. The cross-correlation functions are smoother and have smaller scatter.

PFS will introduce a significant change to Subaru's measurements by making precise spectroscopic observations of previously detected LAEs which will enable more precise computations of their positions, thus allowing 3D cross-correlations. However, this important change in LAE information will not result in significant improvement of the measured cross-correlations. This is due to the fact that the same number of LAEs will describe the observed 3D volume poorer than the 2D surface, increasing the shot noise significantly. Thus, gained information about positions of LAEs and the increase of the shot noise will mostly cancel each other out.

Unlike Subaru's observations which will mostly give insight into cross-correlations towards the end of reionization, since only cross-correlations at $z = 6.6$ are statistically significant, WFIRST will give us insight into higher redshift LAEs and LBGs, with redshift range $7.5 < z < 12$. It is expected that WFIRST will observe spectroscopically around 900 LAEs/deg^2/z at $7.5 \leq z \leq 8.5$. I computed the 21cm-LAE cross-correlation by SKA and WFIRST at $z = 8.06$, which was chosen due to being the only available simulation box in this redshift range. As SKA noise at this redshift is still low and the expected detected LAE density is rather large the resulting cross-correlations follow the theoretical shape with very small differences due to low noise levels. However, the shot noise increases on small scales enabling observations only below $k \sim 0.6\ h\ \mathrm{Mpc}^{-1}$.

In summary, the 21cm-LAE cross-power spectrum is a powerful probe of the EoR which could provide invaluable information on the progress of reionization and the typical scale of ionized regions at different redshifts. Observations with LOFAR and HSC will finally make detection of the 21cm-LAE cross-power spectrum possible at redshift 6.6, as they both plan to observe the ELAIS-N1 field. These observations are going to be very challenging and have substantial problems with noise, but they will still be able to detect the large scales of the cross-power spectrum, which is expected to show an anti-correlation between the two fields. Observations of the cross-correlations between SKA and future galaxy surveys, such as Subaru HSC, Subaru PFS, and WFIRST, will cover a range of different redshifts in reionization. 21cm observations with SKA will have significantly lower noise levels than the observations with LOFAR. WFIRST will observe a large number density of LAEs thus reducing the shot noise too. Both effects will result in more precise cross-correlations and

with increasing the observable scales. Combining the 21cm-LAE cross-correlations between SKA and future galaxy surveys will show the evolution of the cross-correlation across different redshifts, and could give us information about the evolution of the 21cm-LAE interaction, averaged ionized fractions at particular redshift and size of ionized regions.

Chapter 3

Predicting observations of the 21cm power spectra with SKA

21cm power spectra are considered a great probe of the large scales of the high-redshift Universe. In this chapter I will explore the possibility of probing the small scales with 1D LOS 21cm power spectra and asses their observability with SKA at different redshifts throughout the reionization epoch. I will also show 2D 21cm power spectra and explore how their observability with SKA depends on the direction of observation.

3.1 Simulations

To compute the 1D and 2D 21cm power spectra I have used the high-resolution hydrodynamical simulation MasiveBlack-II (MBII) [59], based on P-GADGET [107], which evolves a Λ cold dark matter (ΛCDM) cosmology down to redshift $z = 0$ in a $(100\ h^{-1}\ \mathrm{Mpc})^3$ volume. It has $\sim 10^6\ \mathrm{M}_\odot$ mass resolution and a self-consistent model implemented for cooling, star formation, black holes accretion and associated feedback.

The initial conditions for MBII were generated with the CMBFAST transfer function at $z = 159$. The cosmological parameters used were: $\Omega_\lambda = 0.725$, $\Omega_m = 0.275$, $\Omega_b = 0.046$, $n_s = 0.968$, $\sigma_8 = 0.816$ and $h = 0.701$, which are consistent with the WMAP 7-year data [62]. Haloes were identified with the friends-of-friends procedure [22] which was applied to the dark matter particles with a linking length $b = 0.2$ times the mean interparticle separation. Gas, stars and black holes are associated to their nearest dark matter particles. The subhalo finder SUBFIND [109] was then used on the particles in the identified haloes and computed the local density for each particle. Galaxies are defined as haloes identified with SUBFIND with at least 64 star particles.

For star-forming gas the multi-phase model developed by [108] was adopted. Star formation is given by the Schmidt-Kennicutt Law [58] where the SFR is proportional to the density of the cold clouds ($\rho_{SFR} \propto \rho_{gas}^N$, where $N = 1.5$). Star particles were formed probabilistically from gas particles according to their SFRs. Above a critical density threshold a thermal instability operates producing a two-phase medium consisting of cold

clouds embedded in a low-density gas at pressure equilibrium. Stars form from the cold clouds, and short-lived stars end up as supernovae, supplying the surrounding gas with 10^{51} erg of energy. This heats the diffuse phase of the ISM and evaporates cold clouds, thus establishes self-regulation for star formation. The cloud evaporation process and the cooling function of the gas determine the temperatures and the mass fractions of the 'hot and cold' phases of the ISM. A parametrization of stellar winds is also included. For more details about star formation please see [58].

Black holes are modelled as collisionless sink particles within newly collapsing haloes, which are identified by the friends-of-friends halo finder at regular time intervals. If the halo does not already contain a black hole a seed black hole of mass $M_{seed} = 5\times10^5\ h^{-1}\ \mathrm{M}_\odot$ is inserted into a halo with mass $M_{halo} \geq 5 \times 10^{10}\ h^{-1}\ \mathrm{M}_\odot$. Black holes grow by merging with other black holes or by accreting gas in its surrounding region with an accretion rate $\dot{M}_{BH} = \frac{4\pi G^2 M_{BH}^2 \rho}{(c_s^2+v_{BH}^2)^{3/2}}$, where v_{BH} is the velocity of the black hole relative to the surrounding gas, ρ and c_s are the density and sound speed of the hot and cold phase of the ISM gas. The accretion rate is allowed to be mildly super-Eddington but limited to a maximum allowed value 2 × Eddington rate to prevent artificially high values. The bolometric luminosity of the black hole is proportional to the accretion rate, $L_{bol} = \eta \dot{M}_{BH} c^2$, [104], where $\eta = 0.1$ is the radiative efficiency, and c is the speed of light. The black holes feedback is implemented through the radiated energy, where 5% of it couples thermally to the surrounding gas [25]. It is distributed isotropically on gas particles that are within the black hole kernel (64 nearest neighbours). The feedback energy parameter, $f = 0.05$, is the only free parameter in the model and was set using galaxy merger simulations [25] to match the normalization in the observed MBH-σ relation. For more details about the MBII simulation please check [24, 59]. Simulation outputs were gridded to a 256^3 grid for radiative transfer.

The MBII simulation was subsequently post-processed with the radiative transfer code CRASH [15, 80, 44] to model reionization. CRASH is a 3D radiative transfer code which evaluates the ionization of H and He. The code is based on a combination of ray tracing and Monte Carlo techniques that propagate multi-frequency photon packets through a gas distribution and follow the evolution of gas ionization and temperature over the grid [44]. The CRASH photoionization algorithm includes treatment of the most cosmologically relevant metals in the atomic form: C, O and Si. For more details about CRASH code please check the original CRASH papers. Here we assumed that on average at a particular redshift $f_{esc} = 10\%$ of ionizing photons escapes the galaxies.

From CRASH outputs 21cm brightness temperature maps were computed as

$$\begin{aligned}\delta T_b = 28\mathrm{mK}(1+\delta)x_{HI}\left(1-\frac{T_{CMB}}{T_{spin}}\right)\left(\frac{\Omega_b h^2}{0.0223}\right)\\ \cdot\sqrt{\left(\frac{1+z}{10}\right)\left(\frac{0.24}{\Omega_m}\right)}\left[\frac{H(z)/(1+z}{dv_\parallel/dr_\parallel}\right],\end{aligned} \tag{3.1}$$

where δ is the overdensity, x_{HI} is the mean neutral hydrogen fraction, T_{CMB} is the CMB temperature, T_S is the spin temperature, $dv_\parallel/dr_\parallel$ is the gradient of proper velocity along

the line of sight, while h, Ω_b and Ω_m have their usual cosmological meaning (for more details see [41]). I assume that the gas temperature is equivalent to T_S.

3.2 Power spectra

I define the one-dimensional (1D) power spectrum of the 21cm emission along the LOS of comoving length L as in [29]:

$$P_{21cm}^{LOS}(k_z) = \frac{1}{L}\int \Delta T_b(z)\Delta T_b(z')e^{-ik_z(z-z')}dzdz', \tag{3.2}$$

where $\Delta T_b = \delta T_b - \langle \delta T_b \rangle$ represents fluctuations in the 21cm emission. The dimensionless 1D power spectrum is then defined as

$$\Delta_{1D}^2(k_z) \equiv 2k_z P_b^{LOS}(k_z). \tag{3.3}$$

The two-dimensional (2D) power spectrum of the 21cm emission over an area S is defined as [29]

$$P_{21cm}^{2D} = \frac{1}{S}\langle|\widetilde{\Delta T}_b|^2\rangle, \tag{3.4}$$

where the 2D dimensionless power spectrum is

$$\Delta_{2D}^2(k) \equiv \frac{k^2}{\pi} P_{21cm}^{2D}(k). \tag{3.5}$$

In this chapter I will show dimensionless power spectra and their median values, denoted as $\langle\Delta_{1D}^2(k)\rangle_{median}$ and $\langle\Delta_{2D}^2(k)\rangle_{median}$ for the median of the 1D and 2D power spectra, respectively.

3.3 The SKA noise

The SKA noise is given by the radiometer equation

$$\sigma_{SKA} = \frac{W}{\eta_s}\frac{SEFD}{\sqrt{2N(N-1)\Delta\nu t_{int}}}, \tag{3.6}$$

where W is the weighting factor, η_S is system efficiency, N is the number of stations, $\Delta\nu$ is the bandwidth and $t_{int} = 100$ h is the integration time [54, 64, 12]. The system equivalent flux density is defined as

$$SEFD = \frac{2k_B}{A_{eff}}T_{sys}, \tag{3.7}$$

where k_B is the Boltzmann constant, A_{eff} is the SKA's effective area, and T_{sys} is the system temperature which is defined as

$$T_{sys} = 100 + 400\left(\frac{\nu}{150}\right)^{-2.55}\ \text{[K]}, \tag{3.8}$$

where ν is the observed frequency. The first part of the equation is the contribution from the electronics and the second part is from the sky. For the noise configuration I used the SKA1-LOW configuration as presented in [23]. I will show the 1D dimensionless power spectra of the SKA noise, $\Delta^2_{SKA\ noise}(k_z)$, its average values, $\langle\Delta^2_{SKA\ noise}(k_z)\rangle$, as well as the average values of the 2D dimensionless power spectra of the SKA noise in a direction with one side parallel to the LOS, $\langle\Delta^2_{SKA\ noise_\parallel}(k)\rangle$, and with both sides perpendicular to the LOS, $\langle\Delta^2_{SKA\ noise_\perp}(k)\rangle$.

3.4 Results

In this section I am going to show the resulting 1D and 2D power spectra of the 21cm emission and the SKA noise at various redshifts between $z = 6.5 - 10$.

Fig. 3.1 shows the 1D LOS 21cm power spectra, $\Delta^2_{1D}(k_z)$, of ten random LOSs that go through either neutral or ionized regions at $z = 7.0$ and $\langle x_{HII}\rangle = 0.41$. LOSs that go through the neutral (ionized) regions are chosen as the LOSs with highest (lowest) HI column densities at this redshift. The power spectra show less power on large scales and more power on small scales. However, the power spectra of LOSs that go through ionized regions have decreased power on small scales that are now mostly ionized. The difference in small-scale power between power spectra of LOSs that go through neutral and the LOSs that go through ionized regions is around an order of magnitude. The power spectrum of the SKA noise in LOS direction is stronger on large scales than on small scales. On small scales, the reduction of the SKA noise opens the possibility to observe the 1D LOS 21cm power spectra. The 1D LOS 21cm power spectra of LOSs that go through neutral regions are observable on all but the largest scales, while the LOSs that go through ionized regions will be dominated by the SKA noise, leaving only the smallest scales observable.

Median values of the 1D LOS 21cm power spectra, $\langle\Delta^2_{1D}(k_z)\rangle_{median}$, at different redshifts are shown in Fig. 3.2. The median values are computed from 10 000 random LOSs at each redshift. The median values of the power spectra show a similar trend as in Fig. 3.1, with slightly more power on small scales than on large scales. The strength of the power spectra is also changing with redshift. It is growing with redshift until redshift $z = 9$ and $\langle x_{HII}\rangle = 0.034$ where it reaches saturation. Median values of the power spectra, and thus most of the LOSs at a particular redshift, at all redshifts that are $z \geq 6.5$ and $\langle x_{HII}\rangle \leq 0.68$ are mostly above the SKA noise levels and thus observable.

Median values of the 2D 21cm power spectra are shown in Fig. 3.3 together with the averaged values of the 2D SKA noise power spectra in the direction with one side parallel, $\Delta^2_{SKA\ noise_\parallel}(k)$, and both sides perpendicular to the LOS, $\Delta^2_{SKA\ noise_\perp}(k)$. The strength of the median values of the 2D 21cm power spectra grows from large towards small scales, with the difference in strength between the largest and smallest scales reaching up to an order of magnitude. The strength of the median values of the power spectra grows with the progress of reionization up to redshift $z = 9$ and $\langle x_{HII}\rangle = 0.034$ where it reaches saturation. The average values of the SKA noise power spectra differ significantly depending on the direction of observation. The average values of the SKA noise power spectra in the direction

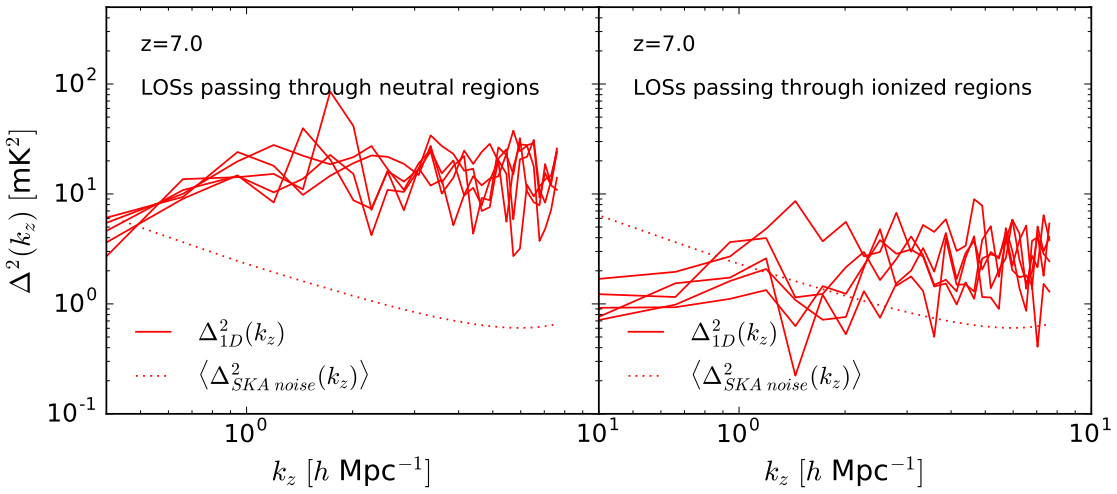

Figure 3.1: 1D LOS 21cm power spectra, $\Delta^2_{1D}(k_z)$, of random LOSs at $z = 7.0$ going through neutral (left panel) and ionized (right) regions. Dotted lines represent average SKA noise power spectrum, $\langle\Delta^2_{SKA\ noise}(k_z)\rangle$, at this redshift.

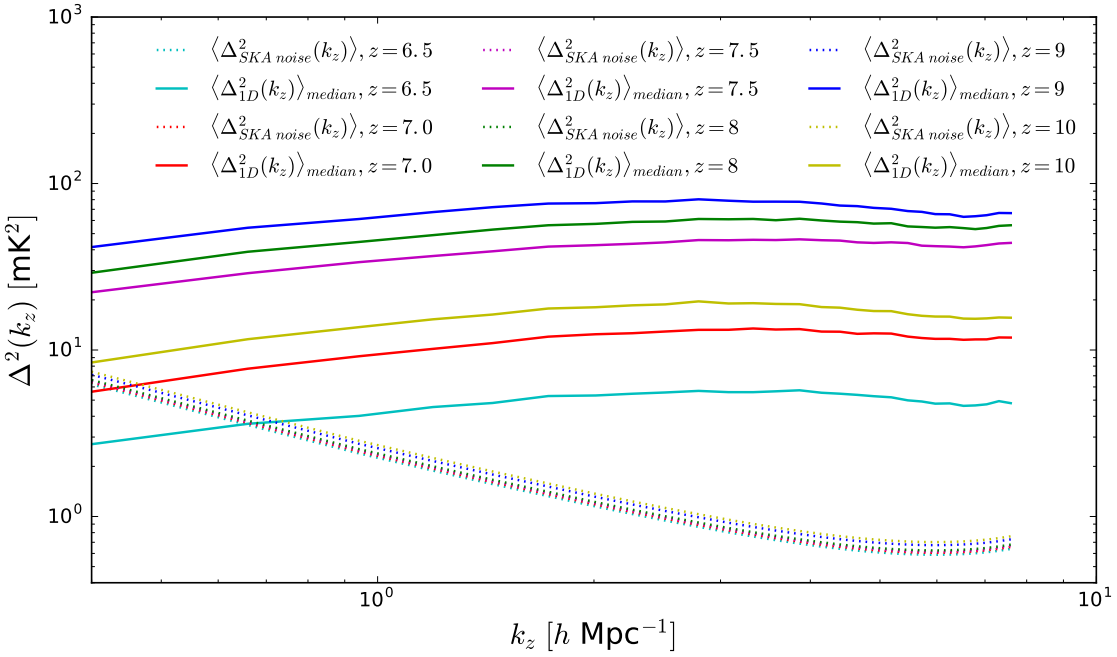

Figure 3.2: Median values of the 1D LOS 21cm power spectra, $\langle\Delta^2_{1D}(k_z)\rangle_{median}$, at different redshifts. Dotted lines represent average SKA noise 1D LOS power spectra, $\langle\Delta^2_{SKA\ noise}(k)\rangle$, at these redshifts.

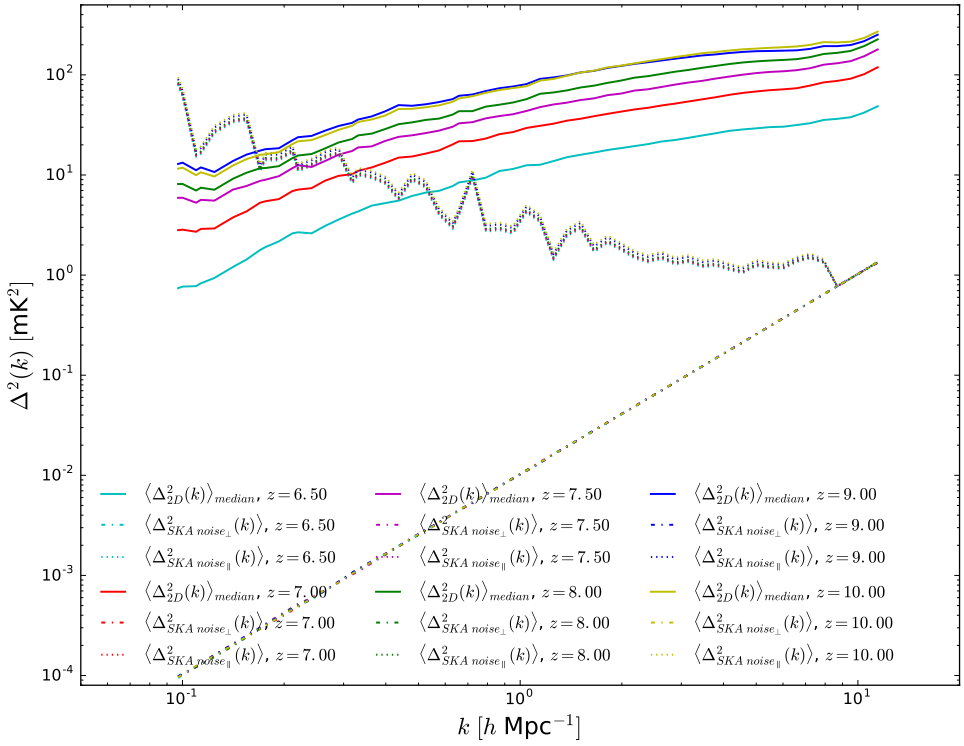

Figure 3.3: Median values of the 2D 21cm power spectra, $\langle\Delta^2_{21cm}(k)\rangle_{median}$, at different redshifts (solid lines). Average SKA noise power spectra at these redshifts are shown in the direction with one side parallel to the LOS (dotted lines), and with both sides perpendicular to the LOS (dashed-dotted).

perpendicular to the LOS, $\langle\Delta^2_{SKA\ noise_\perp}\rangle$ is on large scales orders of magnitude smaller than the signal, it rapidly increases as it progresses towards small scales, however, even on small scales it is still 1-2 orders of magnitude smaller than the signal. The average value of the SKA noise power spectra in the direction with one side parallel to the LOS, $\Delta^2_{SKA\ noise_\parallel}$ is on large scales larger than the signal, but decreases towards small scales, where it is up to an order of magnitude smaller than the signal leaving the scales smaller than $k > 0.2\ h$ Mpc^{-1} at $z = 9$ down to $k > 0.6\ h$ Mpc^{-1} at $z = 6.5$ observable in this direction.

3.5 Discussion and Conclusion

To observe the progress of the reionization process with the 21cm emission, sensitive observations are needed with low noise levels. SKA is a powerful next-generation instrument that will have noise levels low enough to map the high-redshift 21cm emission at redshifts within the Epoch of Reionization and Cosmic Dawn with great precision.

In this chapter I investigated the different directions in which SKA could map the 21cm emission in term of the 2D 21cm power spectra and explored the possibility of probing the small scale emission with 1D LOS power spectra. I show that the 1D LOS power spectrum of the SKA noise has more power on large scales and it reduces towards small scales, making the observations of the 21cm emission on small scales possible at redshifts with average neutral hydrogen fraction $\langle x_{HI} \rangle \geq 0.68$. As expected, the LOSs that go through neutral regions will have a higher power in their power spectra than the LOSs that go through ionized regions. Consequently, the LOSs that go through neutral regions will have a 1D LOS 21cm power spectrum observable at a larger range of scales, with observations possible at scales where $k_z > 0.5\ h\ \mathrm{Mpc}^{-1}$. The LOS that go through ionized regions will instead have observable scales where $k_z > 1 - 2\ h\ \mathrm{Mpc}^{-1}$, depending on the LOS.

I also explored how does the direction of observation influence the 2D 21cm power spectra observations. In the SKA observations noise increases faster with the decreasing frequency than with the increasing angular distance. Thus, the 2D SKA noise power spectra in the FoV where one side is parallel to the LOS will have a lot of power on large scales which decreases towards small scales. This increase of power in the large scales of the noise power spectra will reduce the observability of the large scales, leaving only small scales observable. The 2D SKA noise power spectrum in a FoV with both sides perpendicular to the LOS will, on the other hand, have much less power on large scales and it will increase towards small scales, however, it will always be low enough to make precise observations. Thus, in the direction perpendicular to the LOS SKA will be able to make precise observations of large and small scales at redshifts within the EoR.

Chapter 4

21cm forest

Unlike the 21cm power spectra which would be used to probe the large scales of the Universe, the 21cm forest instead would be a great probe of small scales. In the absorption features in the 21cm forest it is possible to distinguish a range of phenomena that absorb the 21cm line, such as minihaloes and protogalaxies [38, 35, 125], filaments [35], and larger structures like neutral and ionized regions [35, 125]. However, the detection of the 21cm forest is a very challenging endeavour due to the paucity of strong background sources at the redshifts at the beginning or prior reionization.

[10] first explored the observation of the 21cm forest in the quasar spectra. [38, 35] explored the different features of the 21cm forest such as absorption features due to minihaloes, cosmic web and lack thereof due to ionized bubbles. [125] explored the imprint of ionized and neutral regions and their distribution with the equivalent width in the 21cm forest, and more recently [17, 103] explored the observations of the 21cm forest in quasar spectra and [16] the observation in gamma ray burst spectra with LOFAR and SKA. I continue these efforts by exploring the influence of the feedback modeling on the 21cm forest. Radiative feedback could have a great influence on the small scales of the 21cm forest. Thus it is important to explore the 21cm forest with radiation-hydrodynamical simulations that have proper modeling of radiative feedback. In this chapter I will explore the influence of the supernovae feedback on the 21cm forest. I will also explore the influence of the resolution effects on the 21cm forest. To do that I will use a suite of radiation-hydrodynamical simulations of galaxy formation and reionization called Aurora. Aurora simulations have included modelling for the injection of mass and energy by both core-collapse and Type Ia supernovae, and the mass loss due to winds from massive stars and asymptotic giant branch stars. In the simulation suite there are simulations with supernovae feedback included, without it, with differing reionization histories, and different resolutions.

4.1 21cm forest

The 21cm forest is a collection of spectral features due to 21cm absorption in the continuum spectrum of a radio loud source at redshift z_s. As photons with frequency $\nu > \nu_{21cm} =$ 1420.4 MHz redshift to the 21cm line resonance, they get absorbed by neutral hydrogen gas encountered along the line of sight from the source, with an absorption probability of $1 - e^{-\tau_{21cm}}$. τ_{21cm} is the optical depth of the 21cm line and is defined as:

$$\begin{aligned}\tau_{21cm}(z) &= \frac{3}{32\pi}\frac{h_p c^3 A_{21cm}}{k_B \nu_{21cm}^2}\frac{x_{HI} n_H}{T_S(1+z)(dv_\parallel/dr_\parallel)} \\ &= 9.6\cdot 10^{-3} x_{HI}(1+\delta)\left(\frac{1+z}{10}\right)^{3/2}\cdot\left(\frac{T_{CMB}(z)}{T_S(z)}\right)\left[\frac{H(z)/(1+z)}{dv_\parallel/dr_\parallel}\right]\end{aligned} \tag{4.1}$$

where n_H is H number density, δ is the fractional gas overdensity, x_{HI} is the mean neutral hydrogen fraction, T_{CMB} is the CMB temperature, T_S is the gas spin temperature, $A_{21cm} = 2.85\cdot 10^{-15}$ s^{-1} is the Einstein coefficient of the transition, H is the Hubble parameter and $dv_\parallel/dr_\parallel$ is the gradient of the proper velocity along the LOS, which includes both the Hubble expansion and the peculiar velocity ([17, 39]). In this work I assume that the spin temperature is equal to the gas temperature, $T_S = T_{gas}$.

The 21cm forest can present several distinct features: an average optical depth which decreases with the progress of reionization [35]; small scale fluctuations in the optical depth which depend on the density and ionized fraction along the LOS [10, 35]; ionized regions and neutral clumps which leak and absorb the flux respectively [125, 35]; strong narrow absorption lines due to the cosmic web [10, 35] and strong absorption due to minihaloes [38, 35].

The 21cm forest is a great probe of small scale structures, however, it is very difficult to detect it due to a very weak 21cm magnetic hyperfine transition. Another issue is the lack of radio loud sources at the redshifts at the beginning or prior to the EoR as well as the fact that they would be positioned in highly ionized regions which are anti-correlated with high optical depths we are interested in detecting. Observing the 21cm forest would thus be a very challenging endeavour.

4.2 Simulations

I use a suite of radiation-hydrodynamical simulations of galaxy formation and reionization called Aurora [90]. Aurora simulations are based on a modified version of the N-body/TreePM Smoothed Particle Hydrodynamics code GADGET [107].

The simulations have been run from redshift $z = 127$ down to $z = 6$. The initial particle positions and velocities are generated using the Zel'dovich approximation, where particles are arranged along a uniform grid of glass-like structure [120]. The power spectrum of the linear fluctuations is generated by the CAMB code [67]. A ΛCDM cosmological model is adopted with parameters $\Omega_m = 0.265$, $\Omega_b = 0.0448$, $\Omega_\Lambda = 0.735$, $n_s = 0.963$, $\sigma_8 = 0.801$, and $h = 0.71$ [62].

The radiative transfer is implemented using the TRAPHIC code [91]. TRAPHIC solves the time-dependent radiative transfer equation by tracing photon packets at the speed of light and in a photon-conserving manner through the simulation box. It also employs a photon packet merging technique which makes the computational cost of the radiative transfer independent from the number of sources.

Star formation is implemented by using a stochastic recipe of [101] to sample from a Chabrier initial mass function [11] in the range 0.1-100 $M_\odot$ and turn gas particles with densities above $n_H \geq 0.1$ cm^{-3} into star particles, at a rate that reproduces the observed star formation law in simulations of isolated disk galaxies [90].

The injection of mass and energy by both core-collapse and Type Ia supernovae (SNe), and the mass loss due to winds from massive stars and asymptotic giant branch stars are accounted for in the simulations. Core-collapse SNe inject thermal energy with a 30 Myr delay after the creation of a star particle, which approximately corresponds to the maximum lifetime of stars that end as core-collapse SNe. Every core-collapse SN stochastically injects $f_{SN} \times 10^{51}$ erg of thermal energy among a subset of the neighbouring SPH particles, while each Type Ia SN [21] injects and distributes 10^{51} erg of thermal energy among all neighbouring gas particles [102]. Age-dependent injection of energy by Type Ia SNe follows the deterministic method of [102].

Chemical enrichment is also accounted for. This is done by accounting for the age-dependent release of hydrogen, helium, and metals by Type II core-collapse and Type Ia SNe, stellar winds from massive stars, and asymptotic giant branch stars [102, 122]. The total ejected metal mass has been tracked, as well as the ejected mass of 11 individual elements: H, He, C, N, O, Ne, Mg, Si, and Fe.

From the Aurora suite of simulations I used the following simulations:

Simulation with/without SN feedback: has 256^3 dark matter particles and a reionization history with volume averaged neutral hydrogen fraction $\langle x_{HI} \rangle = 0.5$ at $z \sim 7.7$. One of the simulations has SN feedback included, with stellar feedback efficiency $f_{SN} = 0.8$, while the other simulation does not include SN feedback.

Reference simulation: has the best agreement with measurements of the optical depth for electron scattering from the CMB observations with Planck satellite [93]. It has 256^3 dark matter particles and a reionization history with $\langle x_{HI} \rangle = 0.5$ at $z \sim 8.4$. It also has SN feedback included with $f_{SN} = 0.8$.

Higher resolution simulation: is a simulation with 512^3 dark matter particles. This simulation has SN feedback included with $f_{SN} = 0.6$, and it reaches $\langle x_{HI} \rangle = 0.5$ at $z \sim 8.4$.

All the mentioned simulations have the same comoving simulation box size of 12.5 h^{-1}Mpc. From each simulation box 3600 random LOSs were extracted.

The different reionization histories in simulations are accomplished by regulating the stellar feedback efficiency, f_{SN}, and the subresolution escape fraction of ionizing radiation, f_{esc}^{subres}. The subresolution escape fraction only multiplies the stellar luminosities and thus does not fully specify the escape fraction of galaxies in these simulations. Instead, the fraction of ionizing photons that escape a galaxy and are available to ionize the IGM is the product of $f_{esc} = f_{esc}^{subres} f_{esc}^{res}$, where $1 - f_{esc}^{res} \geq 0$ is the fraction of photons absorbed by simulation gas particles on their way to the IGM [90]. The two parameters impact the

SFR functions and the reionization history.

For more details about the simulations I suggest the reader to refer to the original papers [90] and [92].

4.3 Results

In this section I investigate how the prescriptions adopted in the various simulations affect the LOSs in terms of their features and the strength of their power spectra. To do so I will not only compare the same LOS from different simulations, but also statistically quantify the main features in the 21cm forest such as the average optical depth, the size of ionized/neutral regions visible in the forest as leakers/absorbers, and the narrow absorption lines.

I compute the 1D power spectrum of the optical depth along the LOS as [29]

$$P_{\tau}^{LOS}(k_z) = \frac{1}{L}\int \Delta\tau(z)\Delta\tau(z')e^{-ik_z(z-z')}dzdz', \tag{4.2}$$

where L is the length of the LOS and $\Delta\tau = \tau - \langle\tau\rangle$ represents fluctuations of the optical depth in the LOS. I will show median values of dimensionless power spectra, defined as

$$\Delta_{\tau}^2(k_z) \equiv 2k_z P_{\tau}^{LOS}(k_z), \tag{4.3}$$

and median values of the dimensionless power spectra are denoted as $\langle\Delta_{\tau}^2(k_z)\rangle_{median}$. Leakers and absorbers are defined as in [125], i.e. as a continuous part of the spectrum for which

$$\mid \tau - \tau_0 \mid > \tau_{th}, \tag{4.4}$$

where τ_0 is the average optical depth over the whole simulation box, and τ_{th} is the threshold value. Leakers (i.e. ionized regions) are characterized by $\tau - \tau_0 > \tau_{th}$, while absorbers (i.e. neutral clumps) by $\tau_0 - \tau > \tau_{th}$. [125] took the threshold value to be $\tau_{th} = 0.002$ which would show features that could be observed with high signal to noise observation. I, instead, chose $\tau_{th} = 0.0005$ to also show leakers and absorbers at lower redshifts towards the end of reionization where the 21cm signal is not very strong. This choice does not affect the selected leakers and absorbers at higher redshifts significantly.

Narrow absorption lines are defined as regions of ≤ 2 pixels in size (10-15 comoving kpc wide) with $\tau_{nal} > 0.04$ and the optical depth of the surrounding pixels $\tau_{sr} < 0.02$. The optical depths are chosen in such a way to distinguish narrow absorption lines from regular absorbers that are usually bigger in size and with smaller optical depths.

4.3.1 Supernovae feedback

The shape and the strength of absorption features in LOSs depend on many parameters. Here I am going to investigate the effect of SN feedback on the 21cm forest. To do that I will use a pair of simulations, *simulations with/without SN feedback*, where, as the name

suggests, only one of them has SN feedback included. They both have the same reionization history where $f_{SN} = 0.8$ (0.0) in *simulation with (without) SN feedback* and $f_{esc}^{subres} = 0.4$.

Fig. 4.1 shows random pairs of the same LOSs from *simulations with/without SN feedback* at different redshifts. The regions with a lot of wide, connected absorption lines present absorbers (regions rich with neutral hydrogen), and regions with no absorption lines at all present leakers (ionized regions). At higher redshifts one can also see strong narrow absorption lines. I show that higher redshifts are dominated by absorbers with the occasional small leaker, at $z \sim 7.7$ leakers start to dominate until $z = 6$ when the LOSs have no optical depth due to being completely ionized. The same LOSs from *simulations with/without SN feedback* are predominantly equal, however, there are occasional differences in terms of extra leaker/absorber, e.g. at redshifts $\langle z \rangle = 10.20$[1] and $\langle z \rangle = 9.08$ there is a leaker in the LOS from *simulation with SN feedback*, but absorber in the *simulation without SN feedback*. At the LOSs shown at $\langle z \rangle = 7.96$ the opposite happens. These results encouraged the exploration of statistical differences between the LOSs from different simulations.

In Fig. 4.2 I show median values of the 1D LOS power spectra of the optical depth, $\langle \Delta_\tau^2(k_z) \rangle_{median}$, from *simulations with/without SN feedback* at different redshifts. The strength of the $\langle \Delta_\tau^2(k_z) \rangle_{median}$ decreases with the progress of reionization which is due to a reduction in the values of the optical depth. The decrease in strength of the power spectra is bigger on small scales ($k > 60\ h\ \mathrm{Mpc}^{-1}$) which are ionized faster than the large scales. The difference between $\langle \Delta_\tau^2(k_z) \rangle_{median}$ from *simulations with/without SN feedback* is nonexistent at $\langle z \rangle = 11.04$, and it is emerging at $\langle z \rangle = 10.20$ where $\langle \Delta_\tau^2(k_z) \rangle_{median}$ are slightly higher for LOSs from *simulation without SN feedback*, although the difference is barely visible. This dominance of $\langle \Delta_\tau^2(k_z) \rangle_{median}$ from *simulation without SN feedback* is biggest at $\langle z \rangle = 7.96$ and it is more prominent on small scales, although the difference is still relatively small. At $\langle z \rangle = 7.12$ the roles have reversed, where the $\langle \Delta_\tau^2(k_z) \rangle_{median}$ from the *simulation with SN feedback* are now higher, and the difference is stronger than before and it is stronger on all scales.

I show the evolution of the average optical depth, $\langle \tau \rangle$, with the volume averaged neutral hydrogen fraction, $\langle x_{HI} \rangle$, in LOSs from *simulations with/without SN feedback* in Fig. 4.3. As expected, $\langle \tau \rangle$ is decreasing with the progress of reionization. There is no difference between $\langle \tau \rangle$ from *simulations with/without SN feedback* at any redshift. 15% and 85% scatter from the average optical depth also do not show any difference between values from the two simulations.

Fig. 4.4 shows the distribution of leakers at different $\langle x_{HI} \rangle$ in LOSs from *simulations with/without SN feedback*. At the beginning of reionization LOSs from both simulations have leakers that are smaller in size, with most ($\sim 10^5$) leakers at $\langle x_{HI} \rangle = 0.98$ being $r \leq 1\ h^{-1}$ cMpc in length which gradually decreases to 1-2 leakers with $r \approx 7\ h^{-1}$ cMpc. This is because ionized regions at this redshift are still small. With the progress of reionization, and the growth and merging of ionized regions, the maximum length of leakers gradually grows, as well as the number of leakers of all sizes. At $\langle x_{HI} \rangle \approx 0.60$ the first fully ionized

[1]The redshifts denote the redshifts of the simulation box.

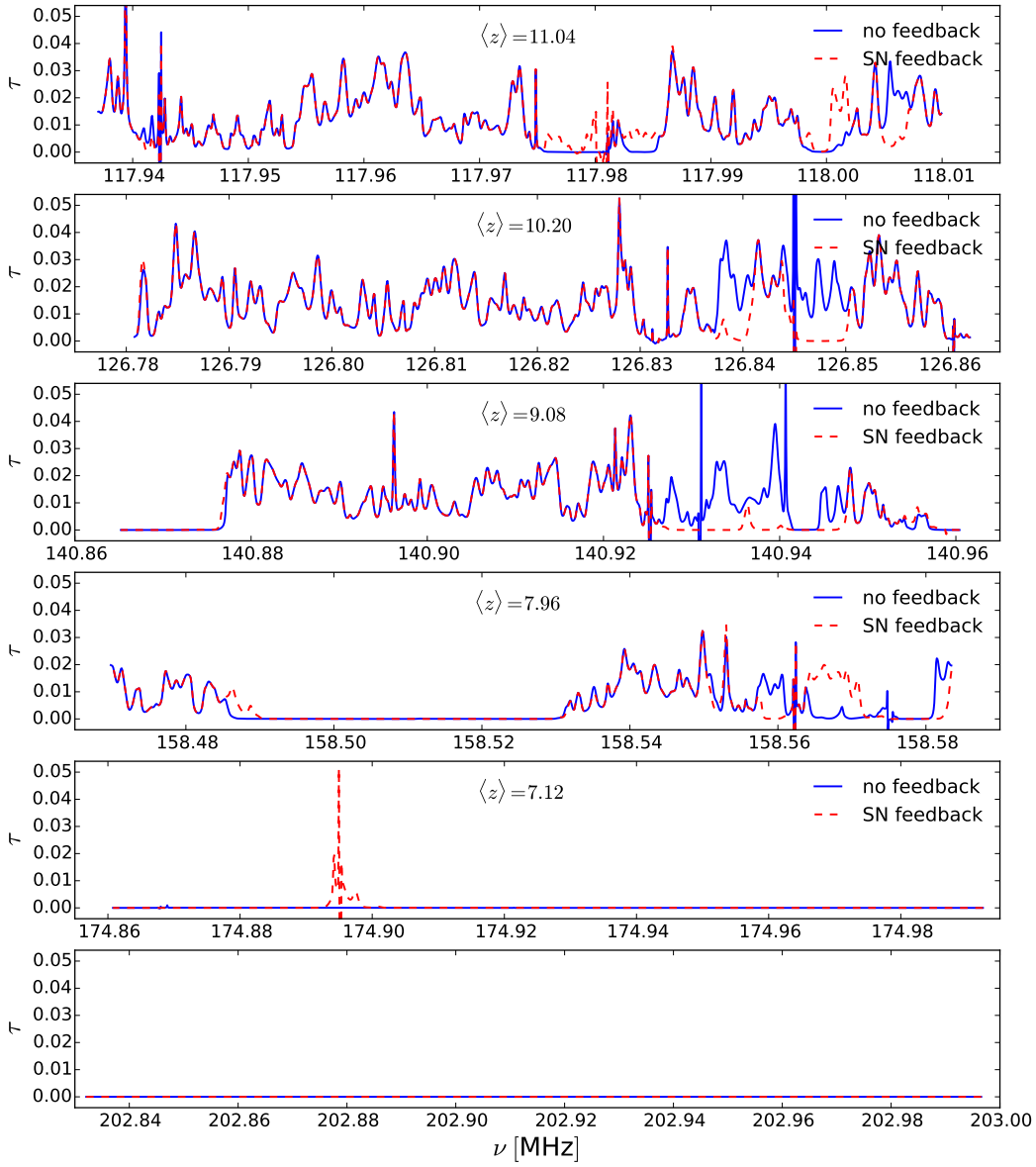

Figure 4.1: Optical depth, τ, of random pairs of LOSs at different frequencies from the *simulation with SN feedback* (red dashed line) and the *simulation without SN feedback* (blue solid). The $\langle z \rangle$ is the average redshift of the simulation box.

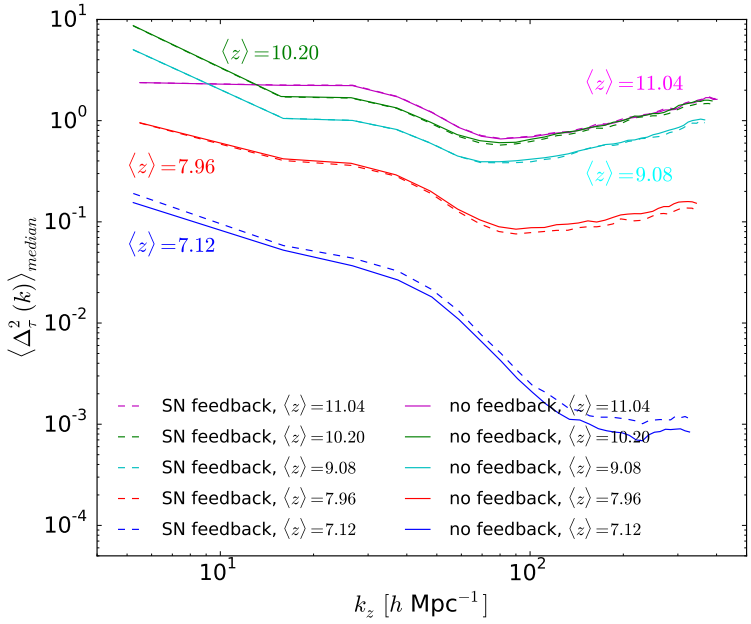

Figure 4.2: Median values of the 1D LOS power spectra of the optical depth, $\langle\Delta^2_\tau(k_z)\rangle_{median}$, at different redshifts from *simulation with SN feedback* (dashed line) and *simulation without SN feedback* (solid).

LOSs emerge, and at $\langle x_{HI}\rangle = 0.24$ ~ 1000 LOSs are completely ionized, with thousands of LOSs that are mostly ionized. However, one should note that the number of leakers is not equivalent to the number of ionized regions, since more LOSs pass through a single ionized region, thus making multiple leaker counts. Comparing the distributions of leakers in LOSs from *simulations with/without SN feedback* it is easy to notice that at $\langle x_{HI}\rangle = 0.98$ there are around half a magnitude more leakers in each size bin in LOSs from *simulation without SN feedback*. At $\langle x_{HI}\rangle = 0.86$ the distributions of leakers from the two simulations starts to even out, and at $\langle x_{HI}\rangle \approx 0.60$ the distributions of leakers are close to equivalent. More towards the end of reionization the situation changes, and at $\langle x_{HI}\rangle = 0.24$ there are slightly more leakers in all size bins in LOSs from *simulations with SN feedback*.

Fig. 4.5 shows the distributions of absorbers at different $\langle x_{HI}\rangle$ in LOSs from *simulations with/without SN feedback*. At the beginning of reionization, where $\langle x_{HI}\rangle \approx 0.98$, absorbers are small and most of them have length $r \leq 1\ h^{-1}$ cMpc which gradually decreases to only a few absorbers with length up to $r \approx 4\ h^{-1}$ cMpc. The length of absorbers slowly increases with the progress of reionization and is reaching up to $r \approx 9\ h^{-1}$ cMpc at $\langle x_{HI}\rangle \approx 0.60$, which reduces to $r \leq 6\ h^{-1}$ cMpc at $\langle x_{HI}\rangle \approx 0.20$. The increase of the number and size of absorbers with the progress of reionization is due to increase in the number of self-shielded regions that remain neutral in the increasingly ionized universe, but also due to the fact that the average optical depth of the LOSs goes down so weaker absorbers satisfy the condition $\tau_0 - \tau > 0.0005$. When comparing the distributions of absorbers at different $\langle x_{HI}\rangle$ in LOSs from *simulations with/without SN feedback* it is noticeable that they are

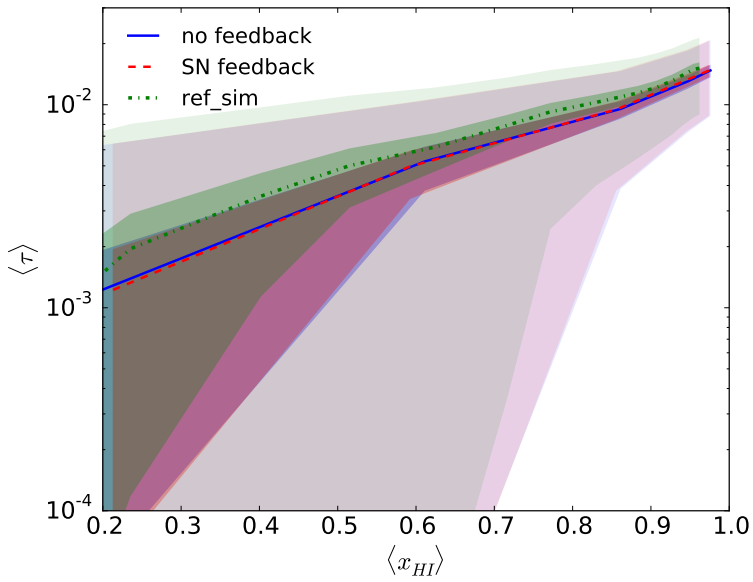

Figure 4.3: Average optical depth, $\langle\tau\rangle$, as a function of the volume averaged neutral hydrogen fraction, $\langle x_{HI}\rangle$, in LOSs from *simulation with SN feedback* (red dashed line), *simulation without SN feedback* (blue solid), and *reference simulation* (green dash-dotted). Shaded areas indicate 15% (darker) and 85% (lighter) scatter from the average value.

following a similar trend as the distributions of leakers. At the beginning of reionization there are around half a magnitude more absorbers in LOSs from *simulation without SN feedback*, then the two distributions start to even out and have almost the same shape and size at $\langle x_{HI}\rangle \approx 0.60$, and towards end of reionization there are slightly more of absorbers in LOSs from *simulation with SN feedback* where the increase in the number of absorbers is stronger for the bigger absorbers with $r = 4 - 6\ h^{-1}$ cMpc.

Fig. 4.6 shows the evolution of the number of narrow absorption lines, n_{nal}, with $\langle x_{HI}\rangle$. As reionization progresses the number of narrow absorption lines decreases, until it reaches zero at the end of reionization. This is because at higher redshifts filaments, which are the main sources of narrow absorption lines, are mostly still neutral and they get ionized with the progress of reionization. Both simulations show similar behaviour in the reduction of the number of narrow absorption lines with the progress of reionization with very little difference, which is due to the fact that both simulations are following the same underlying density distribution.

Considering all the small differences in different absorption features in LOSs from *simulations with/without SN feedback* I suggest they come from the way SN feedback was implemented. Both *simulations with/without SN feedback* have the same f_{esc}^{subres} which impacts SFR functions in such a way to account for the difference in f_{SN} to produce the same reionization history.

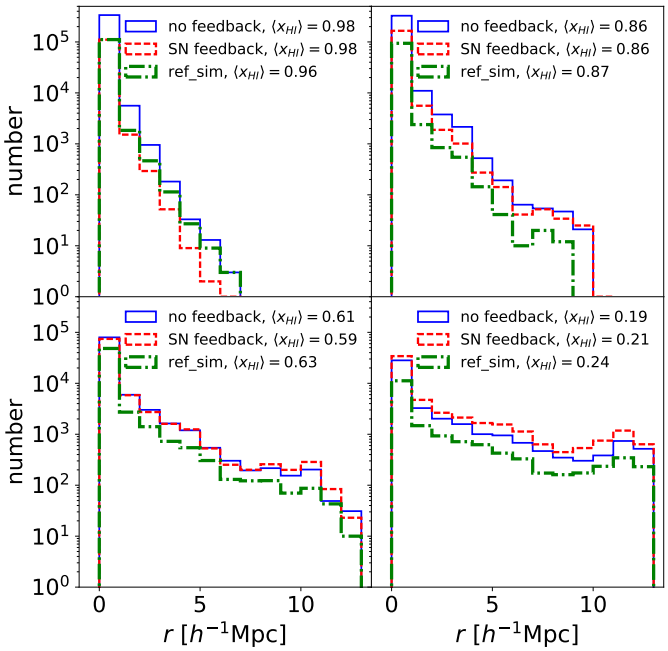

Figure 4.4: Distributions of leakers at different $\langle x_{HI} \rangle$ in LOSs from *simulation with SN feedback* (red dashed), *simulation without SN feedback* (blue solid), and *reference simulation* (green dash-dotted).

4.3.2 Resolution effects

Another parameter that impacts the simulated 21cm forest is the resolution. Here I compare a *higher resolution simulation*, which has 512^3 dark matter particles, to the *reference simulation* with 256^3 dark matter particles. Both simulations have the same reionization history where $f_{SN} = 0.08$ (0.06) and $f_{esc}^{subres} = 0.06$ (0.05) in the *reference simulation* (*higher resolution simulation*).

Fig. 4.7 shows the evolution of $\langle \tau \rangle$ with $\langle x_{HI} \rangle$ in the LOSs from the *higher resolution simulation* and the *reference simulation*. At the beginning of reionization, where $\langle x_{HI} \rangle > 0.8$ the $\langle \tau \rangle$ in LOSs from the *higher resolution simulation* is an order of magnitude higher than in the LOSs from the *reference simulation*. The difference in $\langle \tau \rangle$ slowly decreases with the progress of reionization and at $\langle x_{HI} \rangle = 0.2$ the $\langle \tau \rangle$ in LOSs from the *higher resolution simulation* is only 0.5 order of magnitude higher than in the LOSs from the *reference simulation*. This is due to the fact that the *higher resolution simulation* has better resolution of small dense neutral regions with high optical depth (which are also more numerous at higher redshifts) which significantly increase the $\langle \tau \rangle$.

Figures 4.8 and 4.9 show distributions of leakers and absorbers, respectively, at different $\langle x_{HI} \rangle$ in LOSs from the *higher resolution simulation* and the *reference simulation*. The leakers in LOSs from both simulations follow similar distributions as the ones from *simulations with/without SN feedback*. At the beginning of reionization, where $\langle x_{HI} \rangle \approx 0.95$,

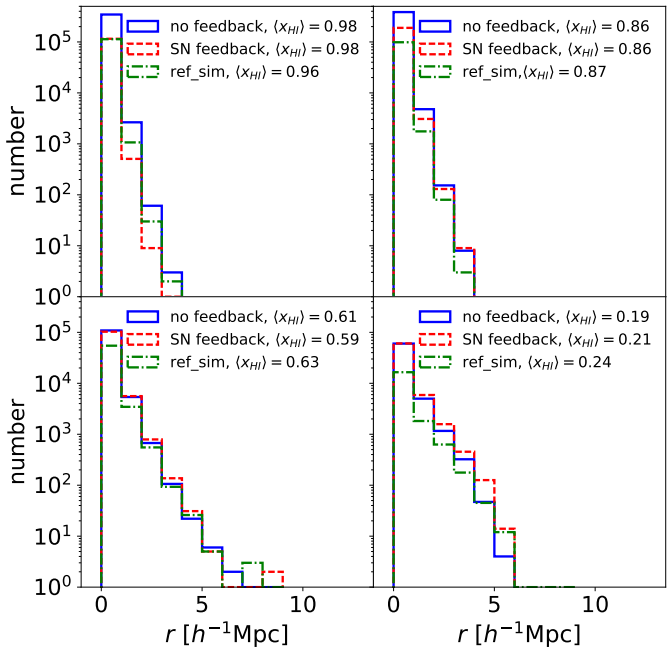

Figure 4.5: Distributions of absorbers at different $\langle x_{HI} \rangle$ in LOSs from *simulation with SN feedback* (red dashed), *simulation without SN feedback* (blue solid), and *reference simulation* (green dash-dotted).

there is ~ 0.1 order of magnitude more of small leakers (up to 1 h^{-1} Mpc) and around half an order of magnitude less in each size bin of bigger leakers in LOSs from the *higher resolution simulation*. They are also smaller in size, with a maximum size of 5 h^{-1} Mpc compared to 7 h^{-1} Mpc in LOSs from the *reference simulation*. These differences decrease with the progress of reionization and at $\langle x_{HI} \rangle \approx 0.23$ the distributions are almost equal. Absorbers follow a similar trend. The distributions from both simulations have a similar shape as the ones from *simulations with/without SN feedback* but there are some differences in the number of absorbers. At the beginning of reionization, where $\langle x_{HI} \rangle \approx 0.95$, there is $\sim 10\%$ more of the smallest absorbers (up to 1 h^{-1} Mpc) and around 1-2 orders of magnitude less of bigger absorbers in the LOSs from the *higher resolution simulation*. They are also smaller in size, with the maximum absorber size in LOSs from the *higher resolution simulation* being 2 h^{-1} Mpc compared to 4 in the *reference simulation*. The same trend continues with the progress of reionization, however, towards the end of reionization at $\langle x_{HI} \rangle \approx 0.23$ the distributions from the two simulations look more equal, but there is still some difference in the number of larger absorbers. The maximum absorber size at $\langle x_{HI} \rangle \approx 0.23$ in the LOSs from the *higher resolution simulation* is 5 h^{-1} Mpc, which is 1 h^{-1} Mpc smaller than in the LOSs from the *reference simulation*. This behaviour is due to the resolution effects in *higher resolution simulation*. Due to a higher resolution of the dense clumps, more photons are needed to ionize a region and as a consequence, the ionized regions are smaller in size. Higher resolution also shows that most of the clumpy

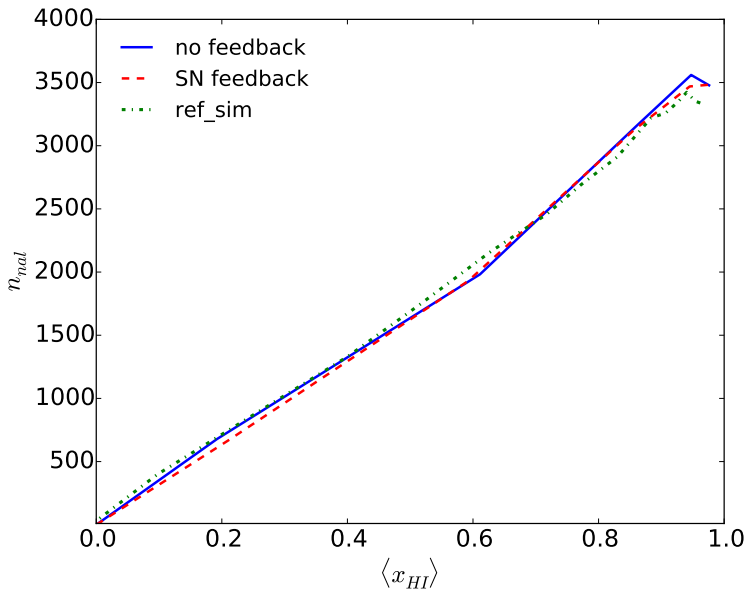

Figure 4.6: The evolution of the number of narrow absorption lines, n_{nal}, with $\langle x_{HI} \rangle$ in *simulation with SN feedback* (red dashed line), *simulation without SN feedback* (blue solid), and *reference simulation* (green dash-dotted).

neutral regions are actually at an average density and only parts are very dense, which decreases the size of the detected absorbers in the LOSs.

The effect of higher resolution becomes more obvious in Fig. 4.10 where I show the evolution of narrow absorption lines, n_{nal}, with $\langle x_{HI} \rangle$. Here I show that there are more than twice as many n_{nal} at all $\langle x_{HI} \rangle$ in the *higher resolution simulation* than in the *reference simulation*. As mentioned before this is due to the better resolution of dense neutral regions and neutral filament structures. This big increase in the number of strong narrow absorption lines is the main reason behind the higher $\langle \tau \rangle$ in the *high resolution simulation*.

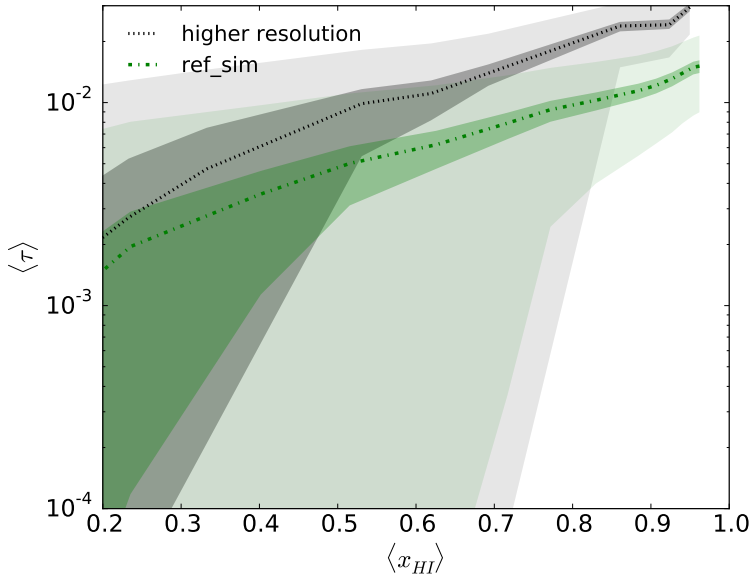

Figure 4.7: Average optical depth, $\langle\tau\rangle$, as a function of $\langle x_{HI}\rangle$, in LOSs from the *higher resolution simulation* (black dotted line) and the *reference simulation* (green dash-dotted). Shaded area indicate 15% (darker) and 85% (lighter) scatter.

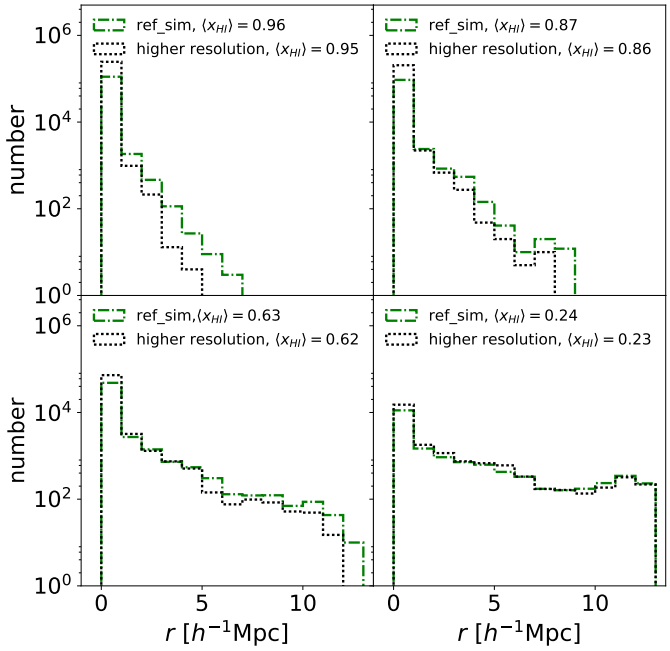

Figure 4.8: Distribution of leakers at different $\langle x_{HI}\rangle$ in LOSs from the *higher resolution simulation* (black dotted) and the *reference simulation* (green dash-dotted).

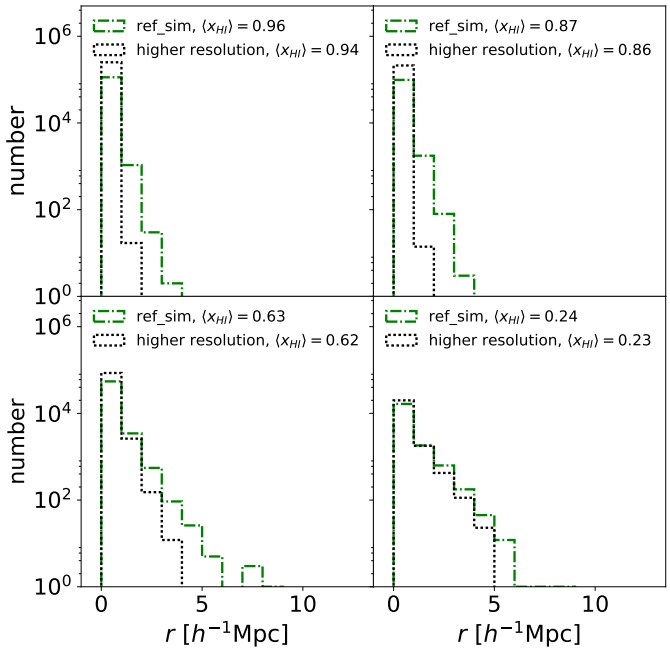

Figure 4.9: Distribution of absorbers at different $\langle x_{HI} \rangle$ in LOSs from the *higher resolution simulation* (black dotted) and the *reference simulation* (green dash-dotted).

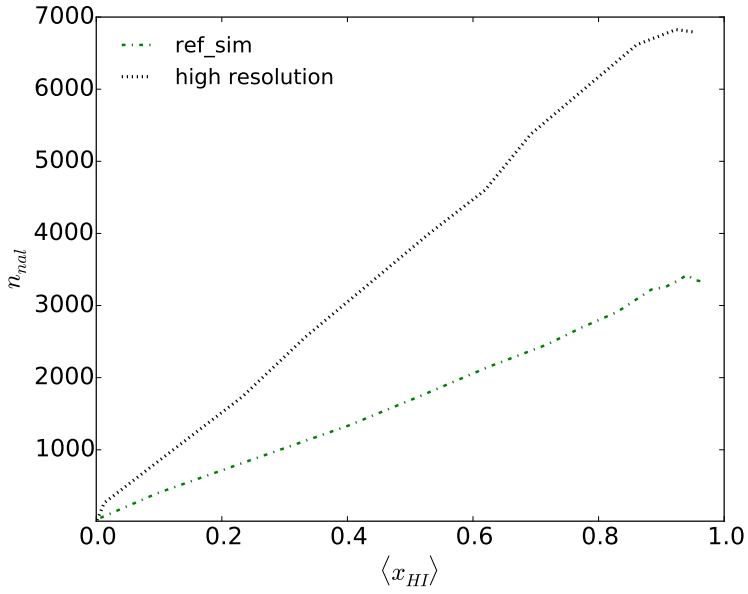

Figure 4.10: The evolution of number of narrow absorption lines, n_{nal}, with $\langle x_{HI} \rangle$ in LOSs from the *higher resolution simulation* (black doted line) and the *reference simulation* (green dash-dotted).

4.4 Observability

In this section I present mock spectra of the 21 cm forest as would be observed with LOFAR and SKA1-LOW.

To design such spectra I assume a background radio source similar to Cygnus A [10, 75], at a redshift z_s, with intrinsic flux density $S_{in}(z_s) = 50$ mJy and power-law index $\alpha_s = 1.05$. I assume that a LOS from the source is going through a region of space with optical depths as previously computed. To these designed spectra I add instrumental effects and noise to mimic observations with LOFAR and SKA1-LOW.

The instrumental response is computed using the pipeline of the LOFAR telescope, as in [17, 16], where I compute visibilities according to the equation for a narrow bandwidth

$$V_\nu(\mathbf{u}) = \sum_{i}^{N_{sources}} I_\nu(\mathbf{s}) e^{-2\pi i \mathbf{u}\cdot\mathbf{s}} + n, \tag{4.5}$$

where $\mathbf{u} = (u, v, w)$ are the coordinates of a given baseline at a given time t, I_ν is the observed source intensity, $\mathbf{s} = (l, m, n)$ represents the direction cosines for a given source direction, and n is additive noise. The noise is defined by the radiometer equation

$$\sigma_n = \frac{W}{\eta_s} \frac{SEFD}{\sqrt{2N(N-1)\Delta\nu t_{int}}}, \tag{4.6}$$

where $W \sim 1.3$ is a weighting factor, $n_s = 0.5$ is the system efficiency, $N = 48$ is the number of LOFAR stations, $t_{int} = 1000$ h is the integration time, and $\Delta\nu = 20$ kHz is the bandwidth. The SEFD is computed as

$$SEFD = \frac{2k_B}{A_{eff}} T_{sys}, \tag{4.7}$$

where k_B is the Boltzmann's constant and A_{eff} is the total collecting area. $T_{sys} = 140 + 60(\nu/150\ \mathrm{MHz})^{-2.55}$ K is the system noise temperature with first contribution being from the electronics and the second from the sky. To produce observational effects for SKA1-LOW I assume that SKA1-LOW will have 1/8 of the LOFAR noise with $\sigma_{n_{SKA}} = 0.1\ \sigma_n$.

In Fig. 4.11 I show the 21cm absorption spectra of a radio source at $z_s = 9.08$ with IGM absorption computed from the *simulation with SN feedback* (upper panel) and *simulation without SN feedback* (middle panel). The LOSs for the simulated absorption spectra, S_{abs}, are chosen from the third panel of Fig. 4.1. The observed spectra, S_{obs}, are modeled for a 1000 h of observation with LOFAR using a 20 kHz bandwidth and a smoothing over a scale $s = 10$ kHz. The spectra show a lot of small scale absorption, where the absorption amplitude goes down to 48.5 mJy. The simulated spectra are identical for the most part with some difference at higher frequencies in terms of two leakers in the LOS from the *simulation with SN feedback* where there are absorbers in the LOS from the *simulation without SN feedback*. The observed spectra are a bit noisy but clearly show the small scale absorption, and is thus easy to recognize the parts of the spectra that are

identical (within the error of the noise) and no absorption features where there are leakers (in the middle panel). Bottom panel shows the signal to noise ratio, S/N, computed as $\mid S_{in} - S_{abs} \mid / \mid S_{obs} - S_{abs} \mid$ which shows a strong detection on most absorption features with $S/N \sim 10$.

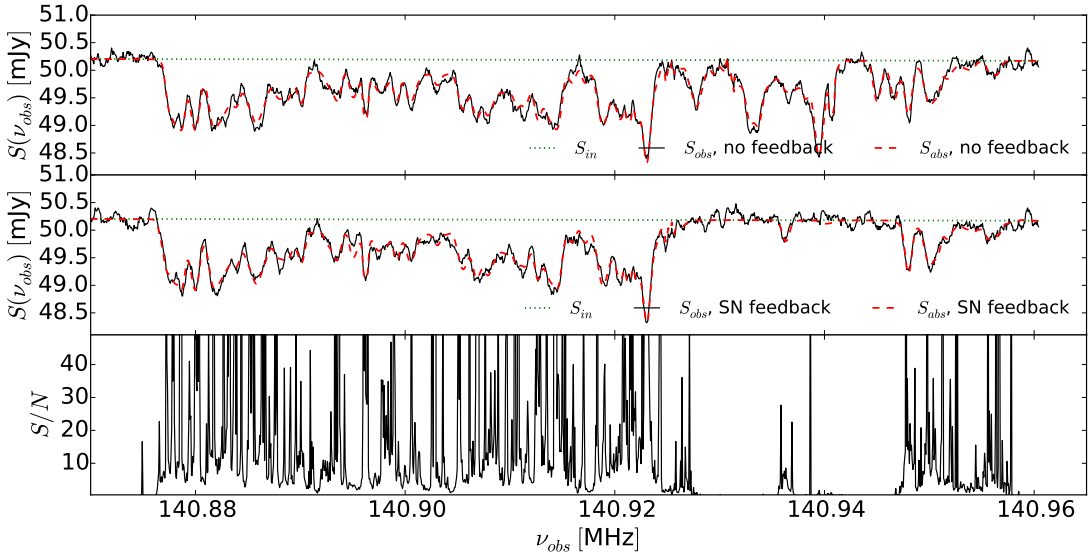

Figure 4.11: Top panel: Spectrum of a radio source at $z_s = 9.08$. The intrinsic spectrum of the radio source, S_{in}, is presented by green dotted line, simulated spectrum of 21cm absorption, S_{abs}, computed from the *simulation without SN feedback*, by red dashed line, and the spectrum of 21cm absorption as would be observed by LOFAR, S_{obs}, by black solid line. Middle panel: The same as in the top panel but with simulated spectrum, S_{abs}, computed from the same LOS from the *simulation with SN feedback*. Both panels show spectra after 1000 h of observation with 20 kHz bandwidth and a smoothing over a scale $s = 10$ kHz. Bottom panel: S/N corresponding to the middle panel.

Fig. 4.12 shows the 21cm absorption spectra of a radio source at $z_s = 11.04$ with the IGM absorption computed from the pair of LOSs from the *reference simulation* (upper panel) and the *higher resolution simulation* (middle panel). The LOSs are the same as in Fig. 4.11, but at an earlier redshift. The observed spectra, S_{obs}, are modeled for a 1000 h of observation with SKA1-LOW using a 20 kHz bandwidth and a smoothing over a scale $s = 10$ kHz. The spectrum designed by a LOS from the *reference simulation* (upper panel) has a lot of small absorption features with the strongest absorption lines reaching down to 48 mJy. The same spectrum designed from the same LOS from the *higher resolution simulation* (middle panel) looks very different and has much stronger features with most features reaching down to 47 mJy and particularly strong one down to 44 mJy. There are also few emission lines from the neutral hydrogen gas. The observed spectrum, S_{obs} is much less noisy here than in Fig. 4.11 due to more precise and less noisy observations

SKA1-LOW will be able to provide. In the bottom panel I show strong signal to noise ratio with $S/N \sim 200$, which is much stronger S/N than in the LOFAR observations.

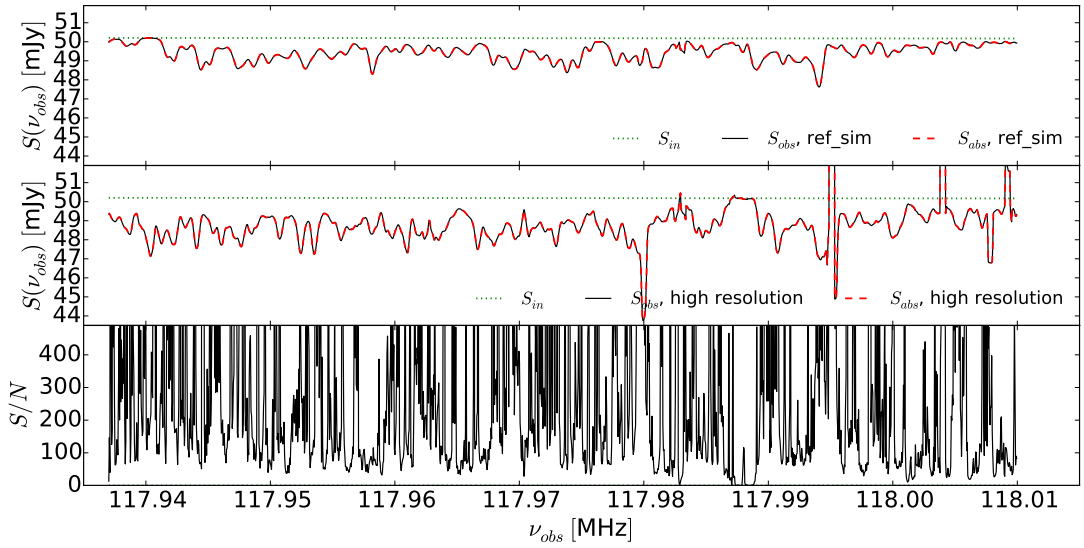

Figure 4.12: Top panel: Spectrum of a radio source at $z_s = 11.04$. The intrinsic spectrum of the radio source, S_{in}, is presented by green dotted line, simulated spectrum of 21cm absorption, S_{abs}, computed from the *reference simulation*, by red dashed line, and the spectrum of 21cm absorption as would be observed by SKA1-LOW, S_{obs}, by black solid line. Middle panel: The same as in the top panel but with simulated spectrum, S_{abs}, computed from the same LOS from the *higher resolution simulation*. Both panels show spectra after 1000 h of observation using a 20 kHz bandwidth and a smoothing over a scale $s = 10$ kHz. Bottom panel: S/N corresponding to the middle panel.

4.5 Discussion and conclusion

In this chapter I focused on exploring the influence of the SN feedback on the 21cm forest. I also explored the influence of the resolution on the 21cm forest.

I use Aurora simulations - a suite of radiation-hydrodynamical simulations of galaxy formation and reionization. These simulations have proper modeling of the injection of mass and energy by core-collapse and Type Ia SNe and the mass loss due to winds from massive stars and asymptotic giant branch stars, making them ideal for these explorations. The simulations also have controlled reionization histories, which are calibrated with two parameters: stellar feedback efficiency, f_{SN}, and the subresolution escape fraction of ionizing radiation f_{esc}^{subres}.

From the Aurora suit of simulations I used *simulations with/without SN feedback*, the *reference simulation*, and the *higher resolution simulation*. To quantify the differences

between LOSs from different simulations I checked the differences between the same LOSs from different simulations and statistically analyzed the different absorption features in the LOSs at particular redshift in terms of 1D LOS power spectra, average optical depth at a particular redshift, the distributions of absorbers and leakers at different redshifts, and the number of narrow absorption lines.

From a comparison between LOSs from *simulation with/without SN feedback* I conclude that while there are some differences in the pairs of LOSs from the two simulations in terms of an extra leaker or absorber, statistically the difference is minimal. At the beginning of reionization LOSs from the *simulation without SN feedback* have slightly more leakers and absorbers and a bit more power in the median values of the 1D LOS power spectra, but in the LOSs from the *simulation with SN feedback* it is visible that they balance out by the end of reionization. These differences do not affect the average optical depths or the number of narrow absorption lines which follow the same behaviour over the whole process of reionization in both simulations. Thus, we found minimal differences between the absorption features in the LOSs from the two simulations which is also due to the fact that the simulations were calibrated to have the same reionization histories.

The resolution effects showed the biggest impact on the 21cm absorption features. The *higher resolution simulation* can better resolve both the dense ionized and dense neutral regions, spending more photons on ionizing the universe and thus resulting in smaller leakers and absorbers. However, it also better resolves small filaments which significantly increases the number of strong narrow absorption lines which as a consequence strongly boosts the average optical depth in the LOSs from this simulation.

I also investigated how the different prescriptions in the simulations affect the observations. To do that I simulated instrumental effects and noise for LOFAR and SKA1-LOW which were added to the simulated spectra from different simulations. I considered a source like Cygnus A [10, 75] with an intrinsic flux density of 50 mJy and a power-law index of 1.05. While LOSs from the *simulations with/without SN feedback* produce spectra that have some differences in terms of an extra leaker/absorber they are otherwise identical, thus the addition of SN feedback does not produce a considerable effect on the observations. However, the LOSs from the *higher resolution simulation* have a significant increase in the structure on small scales compared to the same LOS from the *reference simulation* that follows the same reionization history. Most of these small scale features can be easily observed with LOFAR when smoothing is applied, and all of them could be very precisely observed with SKA1-LOW. SKA1-LOW will have the signal to noise ratio $S/N \sim 200$, which is an order of magnitude higher than in observations with LOFAR.

To conclude, in this chapter I show that simulations with SN feedback prescriptions implemented in such a way to follow the same reionization histories as in the simulations that do not have SN feedback included show minimal statistical differences in the various absorption features. The resolution, on the other hand, has a major influence on the small scales probed by the 21cm forest. They also greatly affect the strength of the observed spectra and have more of stronger absorption features that could be easily observed with LOFAR. SKA will be able to do very precise observations of the 21cm forest and will be able to resolve the smallest features with strong S/N which will be an order of magnitude

stronger than in LOFAR observations. To accurately predict the future 21cm forest observations it is thus imperative to have detailed simulations with feedback prescriptions included and a great resolution.

Conclusion

The Epoch of Reionization is the last great transition of the Universe, a transition from a neutral to an ionized Universe driven by young star-forming galaxies which reionize their surrounding IGM. While our understanding of the processes behind this transition grows rapidly, the observations of the galaxies responsible for it or the 21cm signal from the neutral intergalactic gas are only starting. LOFAR and MWA have started the pioneering attempt of mapping the 21cm signal from the intergalactic neutral hydrogen, and SKA is being built to continue this effort. Similarly, the mapping of the high-redshift LAEs initiated by HSC will be followed by PFS and WFIRST. To gain most from these data it is important to make predictions to asses the best observational strategies and explore possible synergies between different instruments. In this thesis I join these efforts by focusing on three different problems that could help shed more light on this elusive epoch.

I explored the cross-correlation between the 21cm signal and the LAEs observed with LOFAR and Subaru HSC, as well as possible future arrangements with SKA, for observations of the 21cm line, and HSC, PFS, and WFIRST, for LAEs. HSC will observe two fields of interest at redshifts 6.6 and 7.3, while LOFAR will observe at redshift range $6 - 11.4$ [126, 52]. Both telescopes plan to observe the ELAIS-N1 field at redshift 6.6, providing a possibility to detect the 21cm-galaxy cross-power spectrum. Due to the HSC's narrow-band observations, the observed 21cm-LAE cross-power spectrum will be 2D and suffering from projection effects in addition to the large noise from both instruments. These effects will leave small scales unobservable, however a large scale anti-correlation between the 21cm field and the LAEs should be detectable with a cross-correlation coefficient $r_{21,gal} \approx -0.30$ at levels of significance of $p = 0.003$ at $z = 6.6$ and $r_{21,gal} \approx -0.20$ and $p = 0.08$ at $z = 7.3$. This detection would confirm that the observed 21cm signal is indeed from the EoR and that the EoR follows the "inside-out" model of reionization. Observations of the 21cm-LAE cross-correlation functions instead of the cross-power spectra would be inadvisable for the first generation instruments, since the cross-power spectra can separate the noise by its k-modes, leaving most of the noise on small scales only, while the cross-correlation functions will be evenly overwhelmed by noise on all scales. Future observations with the SKA will have a drastic decrease in the noise levels, and thus its observations of the 21cm-LAE cross-power spectra, observed in synergy with HSC, will mostly depend on the shot noise from the HSC survey. This will significantly improve the range of observable scales of the cross-power spectra, from $k < 0.1\ h\ \mathrm{Mpc}^{-1}$ to $k < 0.3\ h\ \mathrm{Mpc}^{-1}$, as well as the detectable strength of the anti-correlation on the large scales of the cross-correlation

coefficient, up to ~ -0.4 at $z = 7.3$ and ~ -0.6 at $z = 6.6$. PFS will make spectroscopic observations of the same Subaru fields of HSC, and as such open up the possibility for an observation of the 3D cross-correlations. This though is not expected to improve upon the 2D cross-correlations due to the much larger shot noise present in 3D observations. The large shot noise will also render the cross-correlation functions unusable. SKA could also observe the 21cm-LAE cross-correlations jointly with WFIRST, which will detect LAEs in the redshift range $7.5-8.5$. Both instruments should have a relatively small noise and thus enable the cross-power spectrum observations at all scales below $k \sim 0.6\ h\ \mathrm{Mpc}^{-1}$. This observation would give us information about the 21cm-LAE cross-correlation at a higher redshift, and thus contribute to our knowledge of the EoR by opening a window into the 21cm-LAE cross-correlation evolution.

Another focus of this thesis was on the observations of the 1D LOS 21cm power-spectra and the 2D 21cm power-spectra with SKA. 1D LOS 21cm power spectra could make small scales accessible to SKA, as its noise decreases on such scales. I showed that indeed most LOSs at redshifts with $\langle x_{HII} \rangle \leq 0.68$ should be observable. It should also be possible to distinguish between LOSs that pass through ionized regions and the LOSs that pass through neutral regions. The latter ones have up to an order of magnitude stronger small scale power, while LOSs that pass through ionized regions will also come much closer to the values of the noise power spectra, making them much noisier. I also explored how does the direction of observation affect the SKA noise levels in 2D observations. SKA noise strongly increases in the frequency direction, making the large scales overwhelmed with noise in the 2D 21cm power spectra observations where one side is parallel to the LOS. When both sides are perpendicular to the LOS only the noise in angular direction is present, which is orders of magnitude smaller than in the frequency direction and few orders of magnitude smaller than the 21cm power spectra. Thus, all 2D 21cm power spectra observations should be done perpendicular to the LOS.

21cm forest in the absorption spectra of a radio-loud background source could be used to observe the small scales of the IGM. I explored how the supernovae feedback and the simulation resolution affect the 21cm forest and what features in the 21cm forest could be observed with LOFAR and SKA. To do that I used Aurora simulations - a suite of radiation-hydrodynamical simulations of galaxy formation and reionization. I showed that, when comparing features in the 21cm forest designed from LOSs from simulations with/without SN feedback, there is not much statistical difference. Increasing the resolution of the simulations has shown instead a substantial difference due to better resolved small neutral regions, which, in addition to producing stronger absorption lines, also change the reionization history since more photons are required to reionize these regions. The most prominent features in the 21cm forest spectra observed in spectra of a radio-loud background source similar to Cygnus A should be visible if observed with LOFAR, while SKA should be able to resolve even the smallest features, with S/N ratio ~ 200.

It is an exciting time for the observations of the high-redshift Universe. The upcoming pioneering observations of the high-redshift 21cm line will open a window into the Epoch of Reionization and Cosmic Down. The first generation instruments, such as LOFAR and MWA, could, with careful calibration and signal processing, observe the large scales of

the Universe in the 21cm signal from high-redshift neutral hydrogen. Cross-correlation of such an observation with LAEs observation would confirm that this signal is indeed from the EoR. Next generation instrument, SKA, will be able to provide us with first images of the Epoch of Reionization and Cosmic Dawn and show us the evolution of the reionization process. Discovery of a high-redshift radio-loud quasar could open a window into the small-scales of the high-redshift Universe. However, to make the most of these exciting observations it is imperative to focus on the understanding of the systematics of these observations and to study the possible observational strategies and synergies with other instruments.

Bibliography

[1] K. Ahn, I.T. Iliev, P.R. Shapiro und C. Srisawat, *MNRAS* **450** (2015), 1486.

[2] R. Barkana und A. Loeb, *ApJ* **609** (2004), 474.

[3] R. Barkana und A. Loeb, *ApJL* **624** (2005), L65.

[4] R. Barkana und A. Loeb, *ApJ* **626** (2005), 1.

[5] G. Bernardi, A.G. de Bruyn, M.A. Brentjens, B. Ciardi, G. Harker, V. Jelić, L.V.E. Koopmans, P. Labropoulos, A. Offringa, V.N. Pandey, J. Schaye, R.M. Thomas, S. Yatawatta und S. Zaroubi, *A&A* **500** (2009), 965.

[6] H.A. Bethe und E.E. Salpeter: *Quantum Mechanics of One- and Two-Electron Atoms*, 1957.

[7] J.S. Bolton, M.G. Haehnelt, S.J. Warren, P.C. Hewett, D.J. Mortlock, B.P. Venemans, R.G. McMahon und C. Simpson, *MNRAS* **416** (2011), L70.

[8] J.D. Bowman, A.E.E. Rogers, R.A. Monsalve, T.J. Mozdzen und N. Mahesh, *Nature* **555** (2018), 67.

[9] C.L. Brogan, E. Fomalont, J. McMullin, S. Myers, N. Radziwill und M. Rupen, *National Radio Astronomy Observatory Newsletter* **113** (2007), 17.

[10] C.L. Carilli, N.Y. Gnedin und F. Owen, *ApJ* **577** (2002), 22.

[11] G. Chabrier, *PASP* **115** (2003), 763.

[12] E. Chapman, F.B. Abdalla, G. Harker, V. Jelić, P. Labropoulos, S. Zaroubi, M.A. Brentjens, A.G. de Bruyn und L.V.E. Koopmans, *MNRAS* **423** (2012), 2518.

[13] E. Chapman, A. Bonaldi, G. Harker, V. Jelic, F.B. Abdalla, G. Bernardi, J. Bobin, F. Dulwich, B. Mort, M. Santos und J.L. Starck, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)* (2015), 5.

[14] B. Ciardi, A. Ferrara, F. Governato und A. Jenkins, *MNRAS* **314** (2000), 611.

[15] B. Ciardi, A. Ferrara, S. Marri und G. Raimondo, *MNRAS* **324** (2001), 381.

[16] B. Ciardi, S. Inoue, F.B. Abdalla, K. Asad und G. Bernardi, *ArXiv e-prints: 1504.07448* (2015).

[17] B. Ciardi, P. Labropoulos, A. Maselli, R. Thomas und S. Zaroubi, *MNRAS* **428** (2013), 1755.

[18] B. Ciardi und P. Madau, *ApJ* **596** (2003), 1.

[19] R.A.C. Croft, J. Miralda-Escudé, Z. Zheng, A. Bolton, K.S. Dawson und J.B. Peterson, *ArXiv e-prints* (2015).

[20] E. Curtis-Lake, R.J. McLure, H.J. Pearce, J.S. Dunlop, M. Cirasuolo, D.P. Stark, O. Almaini, E.J. Bradshaw, R. Chuter, S. Foucaud und W.G. Hartley, *MNRAS* **422** (2012), 1425.

[21] C. Dalla Vecchia und J. Schaye, *MNRAS* **426** (2012), 140.

[22] M. Davis, G. Efstathiou, C.S. Frenk und S.D.M. White, *ApJ* **292** (1985), 371.

[23] P.E. Dewdney, W. Turner, R. Millenaar, R. McCool, J. Lazio und T.J. Cornwell, *SKA-TEL-SKO-DD-001* (2013).

[24] T. Di Matteo, N. Khandai, C. DeGraf, Y. Feng, R.A.C. Croft, J. Lopez und V. Springel, *ApJL* **745** (2012), L29.

[25] T. Di Matteo, V. Springel und L. Hernquist, *Nature* **433** (2005), 604.

[26] M. Dijkstra, *PASA* **31** (2014), 40.

[27] M. Dijkstra, A. Lidz und J.S.B. Wyithe, *MNRAS* **377** (2007), 1175.

[28] M. Dijkstra, J.S.B. Wyithe und Z. Haiman, *MNRAS* **379** (2007), 253.

[29] A. Ewall-Wice, J.S. Dillon, A. Mesinger und J. Hewitt, *MNRAS* **441** (2014), 2476.

[30] X. Fan, M.A. Strauss, R.H. Becker, R.L. White, J.E. Gunn, G.R. Knapp, G.T. Richards, D.P. Schneider, J. Brinkmann und M. Fukugita, *AJ* **132** (2006), 117.

[31] H.A. Feldman, N. Kaiser und J.A. Peacock, *ApJ* **426** (1994), 23.

[32] E.R. Fernandez, S. Zaroubi, I.T. Iliev, G. Mellema und V. Jelić, *MNRAS* **440** (2014), 298.

[33] G.B. Field, *ApJ* **129** (1959), 536.

[34] F. Fontanot, S. Cristiani und E. Vanzella, *MNRAS* **425** (2012), 1413.

[35] S.R. Furlanetto, *MNRAS* **370** (2006), 1867.

[36] S.R. Furlanetto, *MNRAS* **371** (2006), 867.

[37] S.R. Furlanetto und A. Lidz, *ApJ* **660** (2007), 1030.

[38] S.R. Furlanetto und A. Loeb, *ApJ* **579** (2002), 1.

[39] S.R. Furlanetto, S.P. Oh und F.H. Briggs, *PhysRep* **433** (2006), 181.

[40] S.R. Furlanetto, M. Zaldarriaga und L. Hernquist, *ApJ* **613** (2004), 1.

[41] S.R. Furlanetto, M. Zaldarriaga und L. Hernquist, *MNRAS* **365** (2006), 1012.

[42] N.Y. Gnedin, *ApJ* **535** (2000), 530.

[43] N.Y. Gnedin und J.P. Ostriker, *ApJ* **486** (1997), 581.

[44] L. Graziani, A. Maselli und B. Ciardi, *MNRAS* **431** (2013), 722.

[45] Z. Haiman und M. Spaans, *ApJ* **518** (1999), 138.

[46] G. Harker, S. Zaroubi, G. Bernardi, M.A. Brentjens, A.G. de Bruyn, B. Ciardi, V. Jelić, L.V.E. Koopmans, P. Labropoulos, G. Mellema, A. Offringa, V.N. Pandey, A.H. Pawlik, J. Schaye, R.M. Thomas und S. Yatawatta, *MNRAS* **405** (2010), 2492.

[47] G. Harker, S. Zaroubi, G. Bernardi, M.A. Brentjens, A.G. de Bruyn, B. Ciardi, V. Jelić, L.V.E. Koopmans, P. Labropoulos, G. Mellema, A. Offringa, V.N. Pandey, J. Schaye, R.M. Thomas und S. Yatawatta, *MNRAS* **397** (2009), 1138.

[48] I.T. Iliev, G. Mellema, K. Ahn, P.R. Shapiro, Y. Mao und U.L. Pen, *MNRAS* **439** (2014), 725.

[49] I.T. Iliev, G. Mellema, U.L. Pen, H. Merz, P.R. Shapiro und M.A. Alvarez, *MNRAS* **369** (2006), 1625.

[50] I.T. Iliev, G. Mellema, P.R. Shapiro und U.L. Pen, *MNRAS* **376** (2007), 534.

[51] I.T. Iliev, P.R. Shapiro, P. McDonald, G. Mellema und U.L. Pen, *MNRAS* **391** (2008), 63.

[52] V. Jelić, A.G. de Bruyn, M. Mevius, F.B. Abdalla, K.M.B. Asad, G. Bernardi und M.A. Brentjens, *A&A* **568** (2014), A101.

[53] V. Jelić, S. Zaroubi, N. Aghanim, M. Douspis und L.V.E. Koopmans, *MNRAS* **402** (2010), 2279.

[54] V. Jelić, S. Zaroubi, P. Labropoulos, R.M. Thomas, G. Bernardi, M.A. Brentjens, A.G. de Bruyn, B. Ciardi, G. Harker, L.V.E. Koopmans, V.N. Pandey, J. Schaye und S. Yatawatta, *MNRAS* **389** (2008), 1319.

[55] H. Jensen, M. Hayes, I.T. Iliev, P. Laursen, G. Mellema und E. Zackrisson, *MNRAS* **444** (2014), 2114.

[56] H. Jensen, P. Laursen, G. Mellema, I.T. Iliev, J. Sommer-Larsen und P.R. Shapiro, *MNRAS* **428** (2013), 1366.

[57] L. Jiang, E. Egami, M. Mechtley, X. Fan, S.H. Cohen, R.A. Windhorst, R. Davé, K. Finlator, N. Kashikawa, M. Ouchi und K. Shimasaku, *ApJ* **772** (2013), 99.

[58] R.C. Kennicutt, Jr., *ApJ* **344** (1989), 685.

[59] N. Khandai, T. Di Matteo, R. Croft, S. Wilkins, Y. Feng, E. Tucker, C. DeGraf und M.S. Liu, *MNRAS* **450** (2015), 1349.

[60] K. Kohler, N.Y. Gnedin und A.J.S. Hamilton, *ApJ* **657** (2007), 15.

[61] E. Komatsu, J. Dunkley, M.R. Nolta, C.L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D.N. Spergel, M. Halpern, R.S. Hill, A. Kogut, S.S. Meyer, G.S. Tucker, J.L. Weiland, E. Wollack und E.L. Wright, *ApJS* **180** (2009), 330.

[62] E. Komatsu, K.M. Smith, J. Dunkley und C.L. Bennett, *ApJS* **192** (2011), 18.

[63] L. Koopmans, J. Pritchard, G. Mellema, J. Aguirre, K. Ahn und R. Barkana, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)* (2015), 1.

[64] P. Labropoulos, L.V.E. Koopmans, V. Jelic, S. Yatawatta, R.M. Thomas, G. Bernardi, M. Brentjens, G. de Bruyn, B. Ciardi, G. Harker, A. Offringa, V.N. Pandey, J. Schaye und S. Zaroubi, *ArXiv e-prints: 0901.3359* (2009).

[65] P. Laursen, *ArXiv e-prints: 1012.2886* (2010).

[66] P. Laursen, J. Sommer-Larsen und A.O. Razoumov, *ApJ* **728** (2011), 52.

[67] A. Lewis und S. Bridle, *PhRvD* **66** (2002), 103511.

[68] A. Lewis, A. Challinor und A. Lasenby, *ApJ* **538** (2000), 473.

[69] Y. Li, L. Hernquist, B. Robertson, T.J. Cox, P.F. Hopkins, V. Springel, L. Gao, T. Di Matteo, A.R. Zentner, A. Jenkins und N. Yoshida, *ApJ* **665** (2007), 187.

[70] A. Lidz, S.R. Furlanetto, S.P. Oh, J. Aguirre, T.C. Chang, O. Doré und J.R. Pritchard, *ApJ* **741** (2011), 70.

[71] A. Lidz, O. Zahn, S.R. Furlanetto, M. McQuinn, L. Hernquist und M. Zaldarriaga, *ApJ* **690** (2009), 252.

[72] A. Loeb, *JCAP* **3** (2009), 22.

[73] A. Loeb und S.R. Furlanetto: *The First Galaxies in the Universe*, 2013.

[74] A. Loeb und M. Zaldarriaga, *Physical Review Letters* **92** (2004), 211301.

[75] K.J. Mack und J.S.B. Wyithe, *MNRAS* **425** (2012), 2988.

[76] P. Madau, A. Meiksin und M.J. Rees, *ApJ* **475** (1997), 429.

[77] S. Malhotra und J.E. Rhoads, *ApJL* **617** (2004), L5.

[78] S. Malhotra und J.E. Rhoads, *ApJL* **647** (2006), L95.

[79] X.C. Mao, *ApJ* **790** (2014), 148.

[80] A. Maselli, A. Ferrara und B. Ciardi, *MNRAS* **345** (2003), 379.

[81] I.D. McGreer, A. Mesinger und V. D'Odorico, *MNRAS* **447** (2015), 499.

[82] M. McQuinn, L. Hernquist, M. Zaldarriaga und S. Dutta, *MNRAS* **381** (2007), 75.

[83] G. Mellema, I.T. Iliev, M.A. Alvarez und P.R. Shapiro, *New Astronomy* **11** (2006a), 374.

[84] G. Mellema, I.T. Iliev, U.L. Pen und P.R. Shapiro, *MNRAS* **372** (2006), 679.

[85] A. Mesinger und S.R. Furlanetto, *MNRAS* **386** (2008), 1990.

[86] F. Pacucci, A. Mesinger, S. Mineo und A. Ferrara, *MNRAS* **443** (2014), 678.

[87] J. Park, H.S. Kim, J.S.B. Wyithe und C.G. Lacey, *MNRAS* **438** (2014), 2474.

[88] R.B. Partridge und P.J.E. Peebles, *ApJ* **147** (1967), 868.

[89] A.H. Patil, S. Zaroubi, E. Chapman, V. Jelić und Harker, *MNRAS* **443** (2014), 1113.

[90] A.H. Pawlik, A. Rahmati, J. Schaye, M. Jeon und C. Dalla Vecchia, *MNRAS* **466** (2017), 960.

[91] A.H. Pawlik und J. Schaye, *MNRAS* **389** (2008), 651.

[92] A.H. Pawlik, J. Schaye und C. Dalla Vecchia, *MNRAS* **451** (2015), 1586.

[93] Planck Collaboration, P.A.R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A.J. Banday, R.B. Barreiro, J.G. Bartlett und et al., *ArXiv e-prints: 1502.01589* (2015).

[94] J.C. Pober, A.R. Parsons, J.E. Aguirre, Z. Ali, R.F. Bradley, C.L. Carilli, D. DeBoer, M. Dexter, N.E. Gugliucci, D.C. Jacobs, P.J. Klima, D. MacMahon, J. Manley, D.F. Moore, I.I. Stefan und W.P. Walbrugh, *ApJL* **768** (2013), L36.

[95] J.R. Pritchard und S.R. Furlanetto, *MNRAS* **367** (2006), 1057.

[96] J.R. Pritchard und A. Loeb, *PhRvD* **78** (2008), 103511.

[97] J.R. Pritchard und A. Loeb, *Reports on Progress in Physics* **75** (2012), 086901.

[98] B.E. Robertson, R.S. Ellis, J.S. Dunlop, R.J. McLure und D.P. Stark, *Nature* **468** (2010), 49.

[99] M.G. Santos, L. Ferramacho, M.B. Silva, A. Amblard und A. Cooray, *MNRAS* **406** (2010), 2421.

[100] M.R. Santos, *MNRAS* **349** (2004), 1137.

[101] J. Schaye und C. Dalla Vecchia, *MNRAS* **383** (2008), 1210.

[102] J. Schaye, C. Dalla Vecchia, C.M. Booth, R.P.C. Wiersma, T. Theuns, M.R. Haas, S. Bertone, A.R. Duffy, I.G. McCarthy und F. van de Voort, *MNRAS* **402** (2010), 1536.

[103] B. Semelin, *MNRAS* **455** (2016), 962.

[104] N.I. Shakura und R.A. Sunyaev, *A&A* **24** (1973), 337.

[105] M. Silva, M.G. Santos, A. Cooray und Y. Gong, *ApJ* **806** (2015), 209.

[106] E. Sobacchi, A. Mesinger und B. Greig, *MNRAS* **459** (2016), 2741.

[107] V. Springel, *MNRAS* **364** (2005), 1105.

[108] V. Springel und L. Hernquist, *MNRAS* **339** (2003), 289.

[109] V. Springel, S.D.M. White, G. Tormen und G. Kauffmann, *MNRAS* **328** (2001), 726.

[110] D.P. Stark, R.S. Ellis, K. Chiu, M. Ouchi und A. Bunker, *MNRAS* **408** (2010), 1628.

[111] H. Tashiro, N. Aghanim, M. Langer, M. Douspis, S. Zaroubi und V. Jelic, *ArXiv e-prints: 1008.4928* (2010).

[112] G.B. Taylor, C.L. Carilli und R.A. Perley, Herausgeber. *Synthesis Imaging in Radio Astronomy II*, Band 180 von *Astronomical Society of the Pacific Conference Series* (1999).

[113] A.R. Thompson, J.M. Moran und G.W. Swenson, Jr.: *Interferometry and Synthesis in Radio Astronomy, 2nd Edition*, 2001.

[114] K. Toma, T. Sakamoto und P. Mészáros, *ApJ* **731** (2011), 127.

[115] H.Y. Trac und N.Y. Gnedin, *Advanced Science Letters* **4** (2011), 228.

[116] M.P. van Haarlem, M.W. Wise, A.W. Gunst und G. Heald, *A&A* **556** (2013), A2.

[117] E. Visbal und A. Loeb, *JCAP* **11** (2010), 16.

[118] M. Volonteri und N.Y. Gnedin, *ApJ* **703** (2009), 2113.

[119] D. Vrbanec, B. Ciardi, V. Jelić, H. Jensen, S. Zaroubi, E.R. Fernandez, A. Ghosh, I.T. Iliev, K. Kakiichi, L.V.E. Koopmans und G. Mellema, *MNRAS* **457** (2016), 666.

[120] S.D.M. White: *Formation and Evolution of Galaxies. Formation and Evolution of Galaxies*, In *Cosmology and Large Scale Structure*, herausgegeben von R. Schaeffer, J. Silk, M. Spiro und J. Zinn-Justin. (Januar 1996) Seite 349.

[121] R.P.C. Wiersma, B. Ciardi, R.M. Thomas und G.J.A. Harker, *MNRAS* **432** (2013), 2615.

[122] R.P.C. Wiersma, J. Schaye, T. Theuns, C. Dalla Vecchia und L. Tornatore, *MNRAS* **399** (2009), 574.

[123] J.S.B. Wyithe und A. Loeb, *MNRAS* **375** (2007), 1034.

[124] J.S.B. Wyithe und A. Loeb, *MNRAS* **382** (2007), 921.

[125] Y. Xu, X. Chen, Z. Fan, H. Trac und R. Cen, *ApJ* **704** (2009), 1396.

[126] S. Yatawatta, A.G. de Bruyn, M.A. Brentjens, P. Labropoulos, V.N. Pandey, S. Kazemi und Zaroubi, *A&A* **550** (2013), A136.

[127] B. Yue, A. Ferrara, A. Pallottini, S. Gallerani und L. Vallini, *MNRAS* **450** (2015), 3829.

[128] S. Zaroubi, A.G. de Bruyn, G. Harker und R.M. Thomas, *MNRAS* **425** (2012), 2964.

Acknowledgments

This thesis would not be possible without the opportunity given to me by my supervisor Benedetta Ciardi. I would like to thank her for all her support and patience throughout my PhD process. I would also like to thank the LOFAR group for collaboration and for accepting me among them. I am also grateful to Alireza Rahmati for collaboration on the 21cm forest project. The whole PhD experience would be nothing without my partners in crime, Haakon Andresen, Isabella Söldner-Rembold, Jeffrey Chan, Matuš Rybak, Matteo Bugli, Suhail Dhawan and others, to which I would like to thank for their support and all the times we spent together in worry, laughter and gossips. And last but not least, I would like to thank my husband, Tomislav Damjanović, for his boundless support.